DOCUMENT RESUME

ED 080 544 TM 003 049

TITLE	Developments and Assessments in Educational Centers and Laboratories; The Annual Western Regional Conference on Testing Problems (18th, San Francisco, California, May 9, 1969).
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	9 May 69
NOTE	79p.
AVAILABLE FROM	Educational Testing Service, Princeton, New Jersey 08540 ($2.00)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Conference Reports; *Educational Development; Educational Research; Evaluation Criteria; Evaluation Methods; *Learning Laboratories; *Program Evaluation; *Testing Problems

ABSTRACT

The 1969 Western Regional Conference on Testing Problems dealt with developments and assessments in educational centers and laboratories. The following speeches were presented: (1) "Behavioral Objective Specifications in Evaluation: Relevant or Irrelevant" by Marvin C. Alkin; (2) "Approaches to the Validation of Learning Hierarchies" by Margaret C. Wang; (3) "Some Problems with Regard to Research and Development in Higher Education" by Leland L. Medsker; (4) "Educational Research, Educational Development and Evaluation Studies" by John K. Hemphill; and (5) "The Challenge of Multi-Agency Involvement in Development" by Ray Jongeward. A list of conference participants concludes the report. (KM)

ED 080544

TM 003 049

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

WESTERN REGIONAL CONFERENCE ON TESTING PROBLEMS

PROCEEDINGS

1969

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY

Dorothy Urbans

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

EDUCATIONAL TESTING SERVICE

ED 080544

The Eighteenth Annual
Western Regional Conference on
Testing Problems

Developments and Assessments in Educational Centers and Laboratories

May 9, 1969 · Hilton Inn
San Francisco International Airport

Junius A. Davis, *Chairman*

EDUCATIONAL TESTING SERVICE
Princeton, New Jersey · Berkeley, California

Copyright © 1969 by Educational Testing Service

$2.00 per copy

All Rights Reserved

Preface

The theme for the 1969 Western Regional Conference on Testing Problems was Developments and Assessments in Educational Centers and Laboratories.

The program began with Marvin Alkin discussing whether or not the behavioral objective specifications in evaluation were relevant or irrelevant. Margaret Wang continued by outlining the approaches pursued in the validation of learning hierarchies. Leland Medsker completed the morning session by indicating some of the problems in research in higher education. John Hemphill led the afternoon session by directing his attention to educational research and development in relation to evaluation studies followed by Ray Jongeward who described the challenge of multi-agency involvement in the development of a prescribed academic course for use in rural or deprived areas.

JUNIUS A. DAVIS, *Chairman*

Contents

Morning Program

BEHAVIORAL OBJECTIVE SPECIFICATIONS IN EVALUATION: RELEVANT OR IRRELEVANT 1
Marvin C. Alkin, Director
Center for the Study of Evaluation of Instructional Programs
University of California
Los Angeles, California

APPROACHES TO THE VALIDATION OF LEARNING HIERARCHIES 14
Margaret C. Wang, Director
Testing and Evaluation, Primary Education Project
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, Pennsylvania

SOME PROBLEMS WITH REGARD TO RESEARCH AND DEVELOPMENT IN HIGHER EDUCATION 39
Leland L. Medsker, Director
Center for Research and Development in Higher Education
University of California
Berkeley, California

Afternoon Program

EDUCATIONAL RESEARCH, EDUCATIONAL DEVELOPMENT AND EVALUATION STUDIES 50
John K. Hemphill, Director
Far West Laboratory for Educational Development
Berkeley, California

THE CHALLENGE OF MULTI-AGENCY INVOLVEMENT IN DEVELOPMENT 60
Ray Jongeward, Director
Division of Research and Evaluation
Northwest Regional Educational Laboratory
Portland, Oregon

The Eighteenth Annual Western Regional Conference on Testing Problems

The eighteenth annual meeting of the Western Regional Conference on Testing Problems was convened at 9:15 a.m., Friday, May 9, 1969, in the Hilton Inn at the San Francisco International Airport. Junius A. Davis, Director, Southeastern Office, Educational Testing Service, Durham, North Carolina, presided as chairman.

Behavioral Objective Specifications in Evaluation: Relevant or Irrelevant?

MARVIN C. ALKIN

The question posed by the title of this paper, "The Use of Behavioral Objectives in Evaluation: Relevant or Irrelevant?" is not readily answerable. Indeed, there is no single solution to the question. The use of specified behavioral objectives in evaluation is neither relevant nor irrelevant. It is the threefold thesis of this paper that (1) behaviorally stated objectives are of relevance only to certain stages in the evaluation process; (2) even in those stages where it is relevant to state student behavioral objectives, objectives specification alone ceases to be of singular significance with the increasing complexity of the program; and (3) even in relatively noncomplex programs within stages amenable to objectives specification, there is little research evidence showing whether evaluation using specified student behavioral objectives "makes a difference."

The intent of this paper, however, is not to discount completely the value of specifying objectives in the evaluation of instructional programs; to do so would be ludicrous. Alice and the Cheshire Cat probably said it best in *Alice's Adventures in Wonderland*:

"Come, it's pleased so far," thought Alice and she went on. "Would you tell, please, which way I ought to go from here?"
"That depends a great deal on where you want to get to," said the cat.
"I don't much care where . . ." said Alice.
"Then, it doesn't matter which way you go," said the cat.

But behavioral objectives specification is not necessarily a panacea for evaluation problems of all types. While all enterprises should have a goal, these goals are not necessarily always specifiable in student behavioral terms. I would submit also that the continually broadening definition of evaluation has considerably modified views about the need for specification of behavioral objectives.

The last two years have represented an exciting period in the field of evaluation. Indeed, it would not be an overstatement to maintain that evaluation as a field has just begun to assume an identity of its own. I would agree with Egon Guba* that a major failing of evaluation today stems from the lack of an adequate definition. Past definitions have equated it with either: (1) measurement and testing, (2) statements of congruence between performance and objectives, or (3) professional judgment. None of these by itself is really an inclusive enough definition for the multiplicity of activities now regarded as evaluation. During the past year, a consensus has been developing concerning a broader, more comprehensive definition of evaluation. This expanded view takes into consideration the decision-making functions, since an evaluation must be predicated on, and adapted to, the specific problem or situation under analysis.

In view of the fact that there is no definitive statement of evaluation, it would be inappropriate and inaccurate of me to present my definition as "the" generally accepted one. However, in an effort to provide some framework for this paper, I will, somewhat hesitantly, step forward and present my definition of evaluation. *Evaluation is the process of ascertaining the decisions to be made, selecting related information, and collecting and analyzing that information in order to report summary data useful to decision makers in selecting among alternatives.*

The first part of the definition of evaluation presented here deals with ascertaining the decisions to be made. The decision maker, not the evaluator, determines the questions to be asked or the decisions to be made. The task of the evaluator is to determine from the deci-

* Egon Guba, Director, National Institute for the Study of Educational Change, Indiana University, Bloomington, Indiana.

sion maker the decisions for which information is required. The evaluator can and should, however, point out inconsistencies, potential difficulties, or additional data that might modify the decision maker's views on the relevance of certain decisions.

The second task of evaluation deals with the specifications of required information in light of the system's objectives. The specific nature of the information required will differ, of course, depending upon the kind of decision to be made. The task of the evaluator in specifying information requirements includes the development of the research design of the project, and the selection and/or development of instruments designed to provide the information appropriate to the decisions which must be made.

Data collection and analysis are tasks of prime concern to the evaluator. The evaluator will encounter different problems associated with these tasks depending upon the nature of the decisions to be made.

One of the most vital parts of the evaluation process is reporting to the decision maker. Most evaluators often overlook this function, indeed, often consider it a merely pro forma exercise. If the purpose of evaluation is to provide information that will enable decision makers to form judgments about a program or about alternatives, then, the nature and form of the reporting should be appropriate to the problem and to the audience.

STAGES OF EVALUATION

This definition of evaluation carries with it a concern for the decisions to be made. Thus, if we are to understand the evaluation process, it is necessary to categorize educational decision situations. In this classification it would be necessary to examine the nature and kinds of decisions likely to require evaluative data.

I have identified what I consider to be the five stages of an evaluation. Each is designed to provide and report information useful to a decision maker in making judgments. They are (1) systems assessment, (2) planning, (3) program implementation, (4) program improvement, and (5) program certification. I should acknowledge that I have borrowed liberally in the development of these stages from the work of Malcolm Provus* as well as Daniel Stufflebeam.†

* Dr Malcolm Provus, Director, Board of Public Education, Pittsburgh, Pennsylvania.

† Professor Daniel L. Stufflebeam, Director, Evaluation Center, Ohio State University, Columbus, Ohio.

The first area in which evaluation might take place is in the assessment of needs. Needs assessment is a means of determining the educational objectives most appropriate for a particular situation. The needs may be represented as the gap between the goal and the present state of affairs. Thus, the evaluation problem becomes one of assessing the needs of students, of the community and of society in relation to the existing situation. Needs assessment does not refer to specification of process characteristics appropriate for a district, school, or classroom. The needs assessment must be related to the ultimate behavior of clients of one type or another (pupils, parents, community, etc., all are clients of the school). To put it simply, needs assessment must be a statement of objectives in terms of outputs rather than process characteristics of the system.

No doubt it is obvious to you from these examples, as it is to all of those who have been engaged in needs assessment under a Title III program, that the process of deciding purposes of needs assessments, as well as specifying, collecting, analyzing, and reporting information, is quite different from the methodology and techniques usually employed in typical evaluation.

PLANNING

The planning stage in evaluation is concerned with information which will enable the decision maker to select between alternative processes in order to make a judgment as to which process should be introduced into the system in order to fill most efficiently the critical needs which have been previously determined. After the decision maker receives the needs assessment evaluation, he might make a decision as to *the* appropriate means of fulfilling that need. Alternatively, the decision maker might designate several possibilities and ask the evaluator to provide information on the possible impact of each. Thus, in the planning stage, the evaluator provides the data for an evaluation of a program prior to its inception. The task of the evaluator is to look forward to the attainment of goals and to determine the likely goal achievement or outcomes. To repeat this in yet another way, the purpose of an evaluation in the planning stage is to assess the potential relative effectiveness of different courses of action.

It is quite obvious from this discussion that the collection and analysis of data of the type required for this evaluation stage will be quite different from collection and analysis problems for other stages.

The techniques may require both internal and external evaluation criteria. (The most appropriate technique might be informed judgment or other so called soft data.)

The next step in the evaluation process is determining the extent to which the program has been implemented in the manner in which it was described in the design. (A part of the information specification, collection, analysis and reporting process is the specification of the design or procedures by which each of these activities will be accomplished.)

In the case of an existing program, where no known changes have been implemented, the evaluation task for this stage is to determine the degree to which planning descriptions of the program coincide with the actual program and planning descriptions of the students and the context coincide with the actual students and context.

PROGRAM IMPROVEMENT

The evaluator can assume a leadership role in program improvement by providing as much information as possible about the relative success of its parts. In order to perform program improvement evaluation, it is necessary to recognize the basically interventionist role that the evaluator has been asked to play. As the evaluator identifies problems and collects and analyzes information, data are presented immediately to the decision maker in order that changes to improve the operation of the program may be executed within the system. This stage of evaluation has often been overlooked or ignored by the traditional evaluator who has attempted to reproduce the antiseptic sterility of a laboratory in the real world. This approach may make a fine experiment, but it does little to improve a program which is often not in its final form.

PROGRAM CERTIFICATION

Finally, evaluation must provide information to the decision maker that will enable him to make judgments about the instructional program as a whole. This is the "audit" stage of evaluation. The evaluator might attempt to provide information which will enable the decision maker to determine whether the program should be eliminated, modified, retained or expanded.

In this stage, the need for valid and reliable data would generally mandate that the evaluator attempt to apply as rigid a set of controls as possible. The evaluator might use pre and post test designs and

employ sophisticated statistical techniques for analyzing the data whenever possible. Intervention should be avoided in this stage.

USE OF STUDENT BEHAVIORAL OBJECTIVES IN VARIOUS EVALUATION STAGES

I will discuss each of these stages and the categories of decisions related to these stages in order to demonstrate the relevance or lack of relevance of student behavioral objectives for each. A decision is associated with each of the five stages and it is the job of the evaluator to provide the information that will assist the decision maker in selecting between alternatives for that decision. The nature of the decision at each stage, I believe, will demonstrate that information on the achievement of student behavioral objectives is not relevant to some stages and is not the only source of information appropriate for other stages.

In the discussion that follows I do not mean to imply that the evaluator will necessarily participate in each stage of the evaluation. In some instances prior decisions may already have been made and the evaluator may be asked, simply, to provide information for succeeding stages. In other instances the nature of the information to be collected may be relatively simple and the process of information selection, collection and analysis may be internalized by the decision maker and his staff. However, for the sake of clarity, we will assume a hypothetical situation where the evaluator is asked to provide information for decisions at each of the five stages.

The first question facing the evaluator is related to selection of objectives for the system or modification of existing objectives. Thus, depending upon the situation, the decision maker may want information on whether various constituent bodies (i.e., the community) concur with the existing objectives of the system and what changes are needed. It may be appropriate to present information on the potential relevance of alternative objectives in terms of possible future significance.

In a hypothetical situation, a school principal might be faced with budgetary decisions and want to get some insight as to how best to spend money in an incremental budget. He is anxious to spend this in a manner that is likely to be most beneficial to the school in terms of its needs. The evaluator has been asked to provide information about various possible objectives for the system, including some presently stated objectives which may be inadequately met.

Thus, the evaluator may inform the decision maker that a number of behavioral objectives of the system have been considered highly relevant by the community and that the evidence appears to demonstrate that these have been inadequately met. High on this list might be the students' inability to defend themselves against attacks by other students, i.e., trained in "the art of self-defense." The evaluator might also provide information, which would indicate the potential value of selecting "self-defense" as an objective of the system. Thus, the needs assessment evaluation would provide the decision maker with information that would assist him in selecting between alternative objectives. The information is provided by the evaluator, but the relative weighting of the alternatives must be made by the decision maker.

It is obvious from this example that the major source of information provided by the evaluator in this stage is related to students' behavioral objectives for the system. In essence the evaluator provides alternative objectives along with other descriptive information to the decision maker. The student behavioral objectives are of great relevance in this stage of evaluation.

In the planning stage, the evaluator provides information about possible means of achieving the objectives. The question asked by the decision maker is "What process is to be chosen from among a list of alternatives?" The evaluator is not an instructional development expert and ordinarily should not assume the job of developing a program appropriate to the stated objective. However, the decision maker might have narrowed his choice to several alternatives and would like additional information on each of these alternatives.

In the case previously presented, if we assume that the decision maker has selected a behavioral objective related to self-defense instruction and has considered three alternative processes, then the evaluator might provide information related to each of these processes. The information of necessity will be limited in this pre-implementation stage. The evaluator will examine each of the processes in terms of various internal criteria, such as the extent to which the materials purport to achieve the specified objective, the clarity of the materials and the cost of the materials.

In addition, the evaluator may invoke certain external criteria. An examination might be made of the literature related to the use of this process to determine the extent to which it had been found to be successful in similar situations. In the absence of any evidence

related to the use of these materials, the evaluator might choose to use systematically sampled expert judgment about the potential worth of each of the processes considered. Thus, given the information collected and analyzed for this stage, the decision maker would be in a position to make a more rational choice. While it is true that the processes are examined in relation to potentially desired student outputs, the main source of information for the second stage of evaluation is not information on student behavioral objectives but on processes.

The evaluation related to the third stage, program implementation, has as its purpose providing information on whether the process which was selected has been implemented according to plan and whether the context of the situation in terms of the fixed attributes of the program have been described properly in the planning stage. That is: Did the equipment arrive on time? or does the description of the students in the planning stage, which was considered at the time when the process was selected, correspond with the actual situation? It is obvious that in this stage, also, specification of student behavioral objectives is not of critical importance.

In the example that we have been using, let us assume that the decision maker has examined the alternative processes and has decided to introduce a course in shotgun manufacturing to achieve the objective related to "self-defense." One question for the evaluator is: Did the gun barrels arrive on time?

In the fourth stage, program improvement, specified student behavioral objectives are of major importance. In this stage, the evaluator is concerned with determining changes in students and observing students' achievement on a regular basis in order to provide feedback to the decision maker which will be helpful to him in modifying the program. In addition to information related to the achievement of students on certain objective dimensions, the evaluator has as his function within this stage the provision of information relating to the effect of the introduced process upon other processes of the system. Thus, in the example we have been using, the evaluator might note that while students seem to be doing very well in learning to construct shotguns, there appear to be deleterious effects upon teacher-student relationships. Moreover, other students in the school may, for some reason or another, be afraid of those in the experimental program. Finally, the evaluator may note that the general appearance of the school building has suffered. (The walls are pitted,

and many have large gaping holes in them.) On the basis of this information, the decision maker may choose to modify the program, expand it because of the surprisingly good results, or perhaps even delete it immediately.

Let us assume that the program has been allowed to continue and has gone through the program improvement stage to the point where the decision maker is now satisfied with the program and wants to provide a rigid empirical test. At this juncture, the evaluator may be called upon to provide an evaluation related to the program certification function. The evaluator is not being asked to certify the program, but rather to provide information that will allow a decision to be made about certification. As opposed to the previous stages, the role of the evaluator in the area of program certification is noninterventionist. Thus, in the example noted above, the evaluator will attempt to provide information on the decision to the decision maker on the final (or nearly final) outputs of the system in student or other terms as a function of the course in shotgun manufacturing. Again, student behavioral objectives should be considered. The evaluator will also want to provide information on the extent to which students are now better able to defend themselves. There are, however, a number of other outcomes of the systems that were perhaps not anticipated which might well be reported to the decision maker as part of the program certification evaluation. For example, he might note that there has been a considerable increase in the amount of violence in the community and an increase in the number of armed robberies.

I have attempted to demonstrate in the preceding paragraphs that behavioral objectives are of considerable relevance to various stages of evaluation, are of relevance along with other kinds of information in several stages of evaluation, and of little relevance and, indeed perhaps irrelevant in other stages.

In areas traditionally conceived of as evaluation (i.e., program improvement and program certification), there is ordinarily a great need for specifying objectives in behavioral terms as we have just pointed out. But even here I must sound a dissident note. Those advocating the use of behavioral objectives as the main basis for evaluation are usually concerned only with the individual student or, at most, with the classroom as the unit of analysis. The examination of more complex programs often makes it impossible to state behavioral objectives at the outset. One can think of broadscale educational sys-

tems with outcomes that are not clearly definable and where the process of specifying objectives is an iterative one. The complexities of this kind of system are often so great that to speak of objectives in any concrete sense is to mask the real outputs of the system. The outcomes and consequences of all of the many interactions within a system are very great, and are often at considerable variance with the objectives of the system.

Also, the nature of the context at this macro-level of complexity is of considerable significance. While we would maintain that, at the micro-level, the most important element in evaluation is the specification of objectives, in large educational systems the context or nature of the surroundings has tremendous impact on the outcomes of the system. The *Coleman Report*[1] is just one example of a whole line of research which has tended to substantiate this thinking.

Other difficulties in evaluating complex systems involve accurate specification of the instructional treatment. That is to say, often the instructional treatment is neither clean, easily identifiable nor reproducible. It is, instead, a vast array of complex, interactive elements loosely called "instruction."

Thus, we have shown that what is required in this kind of evaluation is not simply a specification of objectives, but, rather, a total examination of a system, with all of the implications that derive from systems theory. A systems evaluation carries with it the necessity for specifying the inputs and outputs of the system, and the understanding that the process of evaluation must be an interactive one in which successive stages produce additional information.

If we think of evaluation as being the process of selecting, collecting, analyzing and providing information for decision makers, then the implications of the data requirements for the evaluation of complex educational systems are readily apparent. In addition to specifying the objectives of the system and the degree to which the system has met these objectives, data must also be provided on other outcomes (unanticipated outcomes, consequences), on the inputs, on accurate descriptions of the alternative processes used, and on the input-output relationships, especially as they relate to the factors which can be considered by the decision maker.

An activity just beginning at the UCLA Research and Development Center is designed to provide answers about the appropriate information necessary for various decisions. The project, the School Evalua-

tion Project, is being directed by Stephen Klein* and myself and is attempting to develop an information system that will help school principals predict student outputs of their schools and make decisions about how to improve these outputs.

The project is uniquely different from most sociopsychological descriptive studies of education in that the orientation focuses on the decisions made by school principals. The project will attempt to determine information requirements (that is, in terms of the definition of "appropriate evaluations") for each of a number of decisions or classes of decisions. It is hoped that the results of this research will provide insights into the relative importance of various kinds of information, including those related to behavioral objectives, for various types of educational decisions.

LACK OF RESEARCH EVIDENCE

Finally, it is imperative to note that even for relatively discrete units of evaluation, there is no definitive evidence that behavioral objectives specification "makes a difference." It has not been substantiated clearly that specifying objectives in behavioral terms for a program modifies the instructional procedures or changes the amount of student learning that takes place. If, from the point of view of the educator, the most relevant considerations are the decisions that will be made as a consequence of the information reporting, then it will be of utmost concern to determine the impact of describing objectives in behavioral terms. There is little evidence to substantiate that such descriptions and the available data relating to them modify the nature of the subsequent judgments by decision makers.

A study by Eva Baker[2] attempted to contrast the effect that behavioral and non-behavioral objectives have on pupil learning and found no significant differences in items directly measuring the objectives or in the transfer items. However, this study dealt with modification of student outputs as a function of using behavioral objectives rather than the impact of such use on decision makers. In a study in which adult students are the decision makers, Blaney and McKie[3] attempted to determine whether knowledge of instructional objectives in an adult education program assists participants in attaining these objectives. The hypothesis is that the group that was

* Dr. Stephen P. Klein, Executive Officer, Elementary School Evaluation Project, Center for the Study of Evaluation, University of California, Los Angeles.

given behaviorally stated objectives would do significantly better than the control group was confirmed. However, in a typical educational situation, one would think of the teacher or another intermediate control agent as the appropriate decision maker rather than the student. A study presently underway by Dr. Eva Baker at the UCLA Research and Development Center will attempt to examine the use of student response data (that is, information related to the achievement of student behavioral objectives) in relation to the subsequent revisions of the instructional material made by the decision maker.

Intuitive feeling, however, would lead to the view that, all things being equal, it is probably better to specify objectives than to not do so at all. With this in mind, and with a deep conviction at the UCLA Center for the Study of Evaluation that the specification of system objectives should be the function of a local decision maker rather than of an external body, the Center is developing a system to help the decision maker determine such selections.

In an attempt to provide local decision makers with behavioral objectives and appropriate test items, we have established an Instructional Objectives Exchange at the Center. The Exchange is under the direction of Rodney W. Skager* and James Popham† and has been established in response to several problems presently existent in the field. These are:

1. The role of the teacher/decision maker as an objectives selector, rather than as an objectives generator
2. The need for test items related to objectives
3. The imminent duplication of efforts in various parts of the United States.

While the Instructional Objectives Exchange project will function as a clearing-house in the area of objectives and items, our prime intended use of the Exchange at the Center goes beyond this. We plan to use some of the material collected in the Exchange in order to study the form and use of behavioral objectives. For example, we want to answer the following questions:

1. Do alternative modes of stating objectives have a relationship to pupil performance?

---

* Professor Rodney W. Skager, Graduate School of Education, University of California, Los Angeles, California.

† Professor W. James Popham, Graduate School of Education, University of California, Los Angeles, California.

2. Does using behavioral objectives as the basis for determining information requirements modify the nature of the ultimate judgments of decision makers?
3. What are the types of decisions made by teachers, administrators and others who have been presented with objective-based data?

We hope that the results of these studies will provide some insights into the relevance of behavioral objectives as a part of evaluation of relatively well defined instructional programs, particularly in the program development and program certification stages.

A RESPONSE

The activities of the UCLA Center of Study for Evaluation are vitally related to the evaluation problems faced by schools and school districts every day. We regard ourselves as a research and development unit whose goal is to make a difference in education. Our activities in conceptualizing evaluation are designed, among other things, to enable us to understand the potential relevance of various procedures in evaluations of different types. Our School Evaluation Project will hopefully provide insights into the information requirements of decision makers. Our activities related to the Instructional Objectives Exchange and Measurement System project will provide evidence as to the form and use of program objectives in decision making in the improvement and certification stages of evaluation.

The kind of mapping of the domain that is exemplified by the three activities named will ultimately allow us to answer in some definitive way whether the need for specification of objectives in evaluation is relevant or irrelevant.

REFERENCES

1. James S. Coleman, et al, *Equality of Educational Opportunity*, Office of Education Reports, OE-38001, National Center for Education Statistics, Washington, D.C., 1966. Printed by U.S. Government Printing Office.
2. Eva Baker, Professor, Graduate School of Education, University of California, Los Angeles; "Effects on Student Achievement of Behavioral and Non-Behavioral Objectives," *Journal of Experimental Education*, Vol. 37, No. 4, pp. 5–8, Summer, 1969.
3. John P. Blaney and Douglas McKie, Division of Continuing Education, University of British Columbia, Vancouver, B.C., Canada, "Knowledge of Conference Objectives and Effect Upon Learning," unpublished.

Approaches to the Validation of Learning Hierarchies

MARGARET C. WANG*

Several independent lines of investigation over the past decade have been focussing on problems of the temporal order in which cognitive behaviors are acquired. Developmental psychologists, particularly those exploring the implications of Piaget's theories of cognitive development, have been interested in demonstrating the existence of regular sequences in the acquisition of concepts and logical operations. At the same time, test and measurement specialists interested in "criterion-referenced testing" have recognized that test batteries based on reliably established acquisition sequences might offer a means of economically estimating performance on a variety of specific behaviors from a relatively small number of test items. Finally, curriculum and instructional designers have been interested in identifying optimal sequences for teaching new skills and concepts. Although these three groups have rather different goals, their concern with sequence in the acquisition of behavior has given them a common interest in the twin problems of generating and validating "behavioral hierarchies"—that is, sets of behaviors which can be shown to be acquired in an invariant sequence, implying that later behaviors are dependent upon, or in some sense "built out of" earlier ones.

* This paper was co-authored by Lauren B. Resnick and Margaret C. Wang. The research reported herein was supported by the Personnel and Training Branch of the Office of Naval Research by a grant from the Project Follow Through of the U.S. Office of Education and by the Learning Research and Development Center supported as a research and development center by funds from the U.S. Office of Education, Department of Health, Education and Welfare. The opinions expressed in this publication do not necessarily reflect the position or policy of the Office of Education and no official endorsement should be inferred.

The developmental psychologist's interest in hierarchies derives largely from a concern for verifying the existence of invariant stages in development, through which all children pass. Hierarchical "stage" theories of development have been proposed by many developmental theorists, of whom the most frequently cited with respect to cognitive development is Piaget.[8,20] Such theories essentially predict the order in which certain behaviors (concepts, intellective and also physical skills) will appear. They do not necessarily imply a "maturational" as opposed to "learning," or organism-environment interaction theory of how such changes occur.[32]

TABLE 1

CROSS SECTIONAL STUDY ANALYSIS
PER CENT OF CONSERVATION RESPONSES FOR MASS, WEIGHT, AND VOLUME AT SUCCESSIVE AGE LEVELS (N = 25 AT EACH AGE LEVEL)[5]

Type of quantity	Age level						
	5	6	7	8	9	10	11
Mass	19	51	70	72	86	94	92
Weight	21	52	51	44	73	89	78
Volume	0	4	0	4	4	19	25

Most studies of developmental sequence have employed cross-sectional designs in which samples of several ages are tested on a set of behaviors. An empirical sequence can then be derived from the percentages of children able to perform the tasks at various ages. An example of data from a cross-sectional study appears in Table 1. The study, by Elkind,[5] examined the ages at which conservation of mass, weight and volume were acquired. Note that the percentage of children conserving mass, mounts sharply at age 7; the same rise in percentage takes place at age 9 for weight; and not at all (up to the age of 11) for volume. These data show a clear order of difficulty among the three tasks and they suggest the hypothesis that each individual child acquires conservation of mass first, then weight and finally volume.

A cross-sectional study, however, cannot directly test the hypothesis that the order of acquisition is invariant for each individual; i.e., that the behaviors are hierarchically organized. Longitudinal studies, in which an initial sample of children are re-examined over a period of years, would permit the testing of hierarchical sequences. However, longitudinal studies are extremely difficult and costly to mount.

Despite general recognition of their value to developmental psychology, relatively few such studies of intellectual development have actually been conducted.*

A few psychologists have seen in scalogram analysis, originally developed by Guttman[17] as a method of scaling responses to attitude questionnaires, a technique that could combine the power of longitudinal studies to examine intra-individual sequence contingencies with the speed and lower cost of cross-sectional studies.[36] These methods have been applied to sequences of behaviors in the areas of haptic perception, logical judgments, moral judgments,[27] number concepts,[36] and classification skills.[19]

TABLE 2
A PERFECT GUTTMAN SCALE

		TEST				
		1	2	3	4	5
Student	1	1	1	1	1	1
	2	1	1	1	1	0
	3	1	1	1	0	0
	4	1	1	0	0	0
	5	1	0	0	0	0
	6	0	0	0	0	0

Like cross-sectional studies, scalogram studies require the administration of a battery of tests presumed to sample behaviors at various points in a linear hierarchy to a group of subjects. Although the age of subjects may vary, age itself is not the independent variable in scalogram studies. Instead, scores on the test battery are examined for "scalability"—the extent which the tests can be arranged in an order such that passing a certain test reliably predicts passage of all tests lower in the scale.* Table 2 shows a hypothetical set of perfectly scaled data. Subjects are listed down the side, tests across the top. Note that once a subject fails a test ("0" indicates failure), he

* One example of a longitudinal study of intellectual development is Piaget's study of his own three children reported in "The Origins of Intelligence of Children."[28]

* The term "test" is used here and throughout this paper to denote a collection of individual items which are presumed to measure the same behavior and for which a single "pass" or "fail" score can be assigned. Thus, "tests" are treated in this research the way "items" were treated in Guttman's original work. An "objective," as used here, is a description of the behavior sampled in a test. It represents an intended outcome of instruction.

fails all subsequent tests. The existence of such a "perfect" scale, or an acceptable approximation to it, is taken to confirm the existence of a behavior hierarchy. While the sequence of acquisition is not observed directly, it is inferred from the fact that individuals who can perform higher level behaviors show evidence of having also learned, or otherwise acquired, all lower level behaviors. The lower level behaviors, in other words, appear to be prerequisites for the higher level ones.

Educational test designers have become interested in scalogram analysis primarily as a means of constructing test batteries for diagnostic or "placement" purposes.[2,6,21] In such testing, the aim is to determine in which specific parts of a curriculum an individual needs instruction rather than to assess a general "level" of performance or to compare individuals or groups. For this purpose, it is often necessary to test large numbers of specific behavioral objectives. This can be an exceedingly complex and time-consuming procedure.

The existence of empirically validated hierarchies can permit substantial economy in placement testing, since subjects who pass a test at the top of a hierarchy can be assumed to be capable of passing all lower level tests. Thus, by testing the top objectives in a number of hierarchies, a student's general "entering level" can be quickly assessed. Subjects who fail the top-level tests in a given hierarchy can then be tested for the lower level objects to determine specific instruction needs.

To learning psychologists and curriculum designers, hierarchies represent a means of sequencing learning tasks in such a way as to maximize transfer from one task to another in order to facilitate the learning of successively more complex behaviors. This means that the requirement of predicting passage of tests lower in the hierarchy is subordinated to the requirement of generating hierarchies in which training on one task has a predictable effect on learning tasks higher in the hierarchy. These two requirements—prediction downward and learning facilitation upward—are closely related. However, they are not necessarily completely correlated. It is theoretically possible for objectives to scale perfectly, but it is also possible for instruction in a task lower on the scale not to produce significant amounts of transfer to higher level objectives. On the other hand, it may be possible to construct highly efficient instructional sequences which introduce objectives without having first established all prerequisite behaviors specified in a scale. Researchers interested in the use of hierarchies

as a means of sequencing instructional objectives, therefore, are necessarily concerned that hierarchy validation studies seek to establish independently the scaling properties of hierarchies and their learning transfer properties. The extent to which transfer and scaling relationships coincide can then become a matter for empirical investigation.

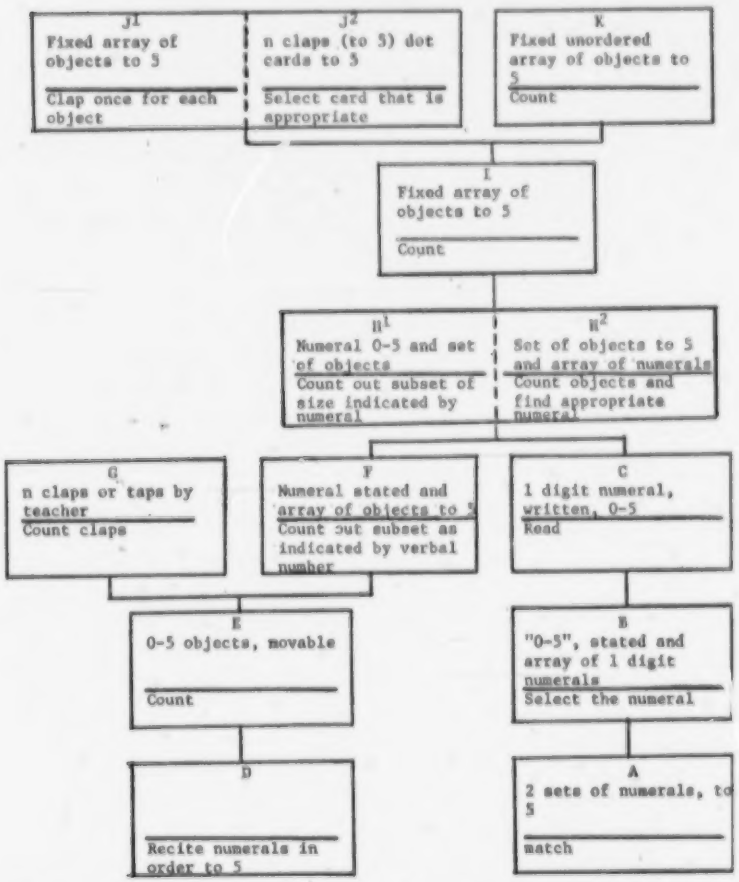
Gagné[10] was the first to formally propose the use of learning hierarchies in designing educational programs, although various methods of "task analysis," leading to hierarchy-like structures, had been used in developing industrial and military training programs for some time.[25] Gagné has outlined a procedure by which behaviors can be analyzed by asking the single question, "What kind of capability would an individual have to possess to be able to perform this task successfully, were we to give him only instructions?" One or more subordinate tasks are specified in response to this question. The question is then applied to the subordinate tasks themselves, and so on successively down the hierarchy until tasks that can be reasonably assumed in the student population are reached. In our own work we have been developing rather more formal methods of generating hierarchies.[29] Our method is based on an analysis of skilled performance that has certain features in common with the technique of "protocol analysis" developed by Newell[26] in connection with information processing and computer-simulation studies. We also insist on a rigorous specification of stimulus and response in our task definitions, which has the effect of keeping each of our tasks more "unitary" than most of Gagné's. Operationally, this means that fewer test items would be needed to sample each task in our hierarchies than in Gagné's.

Figure 1 is an example of one of our hypothesized learning hierarchies. Each box defines a task. The entry above the line defines the stimulus situation; the entry below the line, the response. The simpler behaviors, according to our analysis, appear at the bottom of the chart; the more complex behaviors toward the top. Note that this hierarchy, like most of Gagné's, is non-linear. For example, behavior E is considered prerequisite both to G and F, and H is shown as having two prerequisites, C and F. For instructional purposes, sequences, ABC and DEF could be taught simultaneously, or either one might come first; but both would have to be learned before H could be acquired. This branching characteristic permits us to recognize within a hierarchical framework much of the variety and complexity that characterizes learning patterns. For this reason, we

believe that hierarchies of this kind more accurately reflect psychological reality than do the linear hierarchies mainly used by developmental psychologists[30,19] and by testers.[2] However, a branching hierarchy poses certain knotty problems in validation methodology.

FIGURE 1

HYPOTHESIZED HIERARCHY FOR QUANTIFICATION I



These are the problems to which much of our current work in hierarchies is addressed, and to a discussion of which we now turn.

Our first validation studies were concerned with the "scaling" prop-

erties of a set of hierarchies in the area of early quantification skills. Figure 1 represents one of the hierarchies studied. A battery of criterion-referenced tests* was developed,[37] one for each of the objectives included in the hierarchies. The battery of tests was administered to a random sample of kindergarten children in September, 1968, before any formal instruction in the curriculum was given. The results of these tests were then analyzed for scaling properties.

Our first analyses represented an attempt to adapt existing linear scaling procedures to the validation of branching hierarchies. For this purpose we used the Multiple Scalogram Analysis, a procedure developed by Lingoes.[22] This procedure was selected for several reasons. First, it can not only validate or refute a hypothesized sequence but can also suggest a more optimum sequence or set of sequences. It also provides multi-dimensional information about the tests in a given scale. When the data demand it, it can yield multiple scales rather than rejecting the scale hypothesis for the set treated as a whole. With respect to statistical reliability, MSA contains a measure to control for spuriously high estimates of "reproducibility"—Guttman's classical measure of scalability. This is an important feature of the program, since the possibility of inflated reproducibility indices, due to extreme pass or fail rates on certain tests in the battery, has been one of the major criticisms of Guttman's method in the past.[23,7,16,35,3,4,32] Finally, a computer program has been developed for MSA—the *Format Free Multi-Scaling Program* (SCALE); therefore, MSA is an economical and convenient procedure to use, especially when dealing with large sets of data.

Although the MSA program is capable of picking out multiple scales, these scales are independent of one another, having no objectives in common. Once an objective is selected for inclusion in a scale, it is no longer considered for membership in other scales. For example, with respect to Figure 1, if objective H were to scale with C, B, and A it could not appear in a scale with F, E, and D in the same analysis. Therefore, in order to apply the program to validate a branching hierarchy, it was necessary to test separately each of the linear pathways implied by the hierarchy. For the hierarchy

* "Criterion-referenced test" is an achievement test developed to assess the presence or absence of a specific criterion behavior described in an instructional objective. Such a test provides information about the competence of a student that is independent of the performance of other students. For further discussion of criterion-referenced tests see Glaser.[15]

TABLE 3

A Comparison of the Hypothesized and the Empirical Scales for Quantification I

Analysis 1		Analysis 2		Analysis 3		Analysis 4		Analysis 5	
Hypothesized Scale	Empirical Scale	Hypothesized Scale	Empirical Scale	Hypothesized Scale	Empirical Scale	Hypothesized Scale	Empirical Scale	Hypothesized Scale	Empirical Scale
A	A	A	A	D	D	D	D	D	D
B	K	B	I	E	E	E	E	E	E
C	I	C	B	F	K	F	F	G*	
H_1	B	H_1	C	H_1	I	H_1	H_1		
H_2	C	H_2	H_1	H_2	F	H_2	J_1		
I	H_1	I	J_1	I	H_1	J_1	H_2		
K	H_2	J_1	H_2	K	H_2	J_2	J_2		
		J_2	J_2						
Reproducibility	.926		.908		.926		.904		

* Eliminated from consideration because all S's failed.

shown in Figure 1 we ran five separate analyses: A B C H_1 H_2 I K; A B C H_1 H_2 I J_1 J_2; D E F H_1 H_2 I K; D E F H_1 H_2 J_1 J_2; and D E G.

The input data for the analyses consisted of a pass or fail score for each subject on each test. The index of the degree to which the objectives are sequenced is operationally defined as the reproducibility criterion for Guttman scales:

$$\text{Rep.} = 1 - \frac{\text{Sum of errors}}{\text{Total Responses}}$$

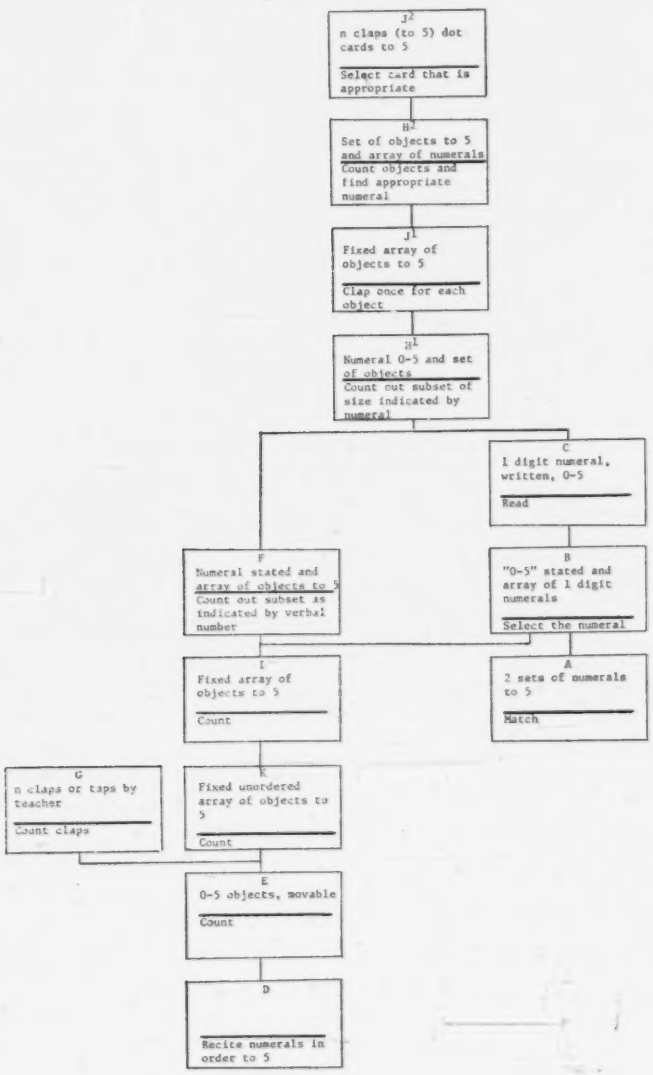
Error is defined as a case where a subject passes a higher level objective and fails a lower objective. In this study, the criterion of reproducibility was set at .85. This meant that only those tests that could enter a scale with a reproducibility equal to or greater than .85 were included in the scale.

The results of these analyses are shown in Table 3. For each analysis the first column shows the hypothesized scale and the second column shows the empirical scale generated by MSA. Analysis 1 shows that K and I (counting ordered and unordered arrays of objects) had been placed too high in the hypothesized sequence. These counting tasks, according to the data, should come before tasks involving numerals (B,C,H_1,H_2). The basic sequence with respect to learning numerals (A, then B, then C, then H), however, was confirmed. Matching numerals (A) appeared as prerequisite to counting, but this may have been an artifact of the very high rate of passing test A. Where nearly all subjects in a sample can perform a behavior, scaling may show it as prerequisite even to unrelated behaviors. Analysis 3 tests the sequence of all counting objectives (D,E,F,I and K) and suggests that counting fixed arrays (K and I) comes before counting out a subset from a larger set (F). Even counting out a set (F), however, should come before using numerals (H_1 and H_2), according to this analysis. In combination, Analyses 1 and 3 suggest that our initial hierarchy introduced numerals too early in the counting sequence. The implication—not directly tested in these analyses—is that counting of various kinds must be established before numeral recognition can be learned. Analyses 2 and 4 support this interpretation, and also suggest a reordering of the sub-objectives in H and J.

On the basis of these analyses, it was possible to construct a new learning hierarchy, rearranging the original objectives. This hierarchy is shown in Figure 2. The five objectives involving counting of

FIGURE 2

REORDERED HIERARCHY FOR QUANTIFICATION



objects (D,E,K,I,F) are now in a linear order, with numeral identification (B) appearing as an upward branch from I. Visual matching of numerals (A) is shown as prerequisite only to numeral identification and reading (B and C) because, despite its apparent relationship to K and I in the empirical scales, it did not seem reasonable to expect that learning visual matching of numerals would help in learning to count. H and J sub-objectives appear in the new order suggested by the analyses. This order seems quite reasonable since both H_1 and J_1 involve counting a set (of objects or events) in response to a symbolic presentation, and both H_2 and J_2 involve selecting symbols to match sets. Counting claps (G) is retained as a separate branch. As with all post-hoc interpretations, of course, it will be necessary to test this reordered hierarchy using new samples of subjects before accepting its validity.

TABLE 4

COMPARISON OF HYPOTHESIZED AND EMPIRICAL SCALE FOR COUNTING OBJECTS AND COMPARISON OF SETS
(N = 37)

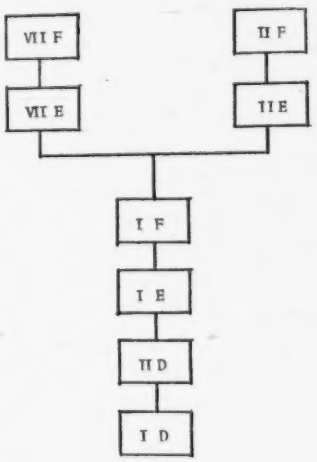
Hypothesized Scale	Empirical Scale		
	Scale 1	Scale 2	Scale 3
Q I D (Rote count 0–5)	I D	VII B	VII D
E (Count moveable objects 0–5)	II D	II E	
F (Count out a set 0–5)	I E	VII F	
Q II D (Rote count 6–10)	I F	VII C	
E (Count moveable objects 6–10)	VII E	II F	
F (Count out a set 6–10)			
Q VII B (Pair sets—equal, unequal)			
C (Pair sets—more, less)			
D (Pair sets—most, least)			
E (Count sets—equal, unequal)			
F (Count sets—more, less)			
*G (Count sets—most, least)			
Reproducibility	.950	.886	1.000

* Eliminated from consideration because all S's failed.

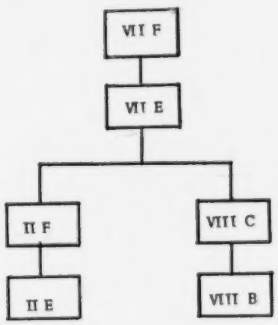
In this first application, Multiple Scalogram Analysis proved usable, although awkward in requiring so many separate analyses. Our next attempt to apply MSA, however, was to reveal more serious complications. Table 4 shows the results of an attempt to test the hierarchical relations between counting skills (Q I and Q II) and two methods of comparing set size, (a) by one-to-one correspondence

(Q VII B,C,D) and (b) by counting each set (Q VII E,F,G). Our hypothesis in this case was a linear one. We predicted that children would first learn to count five objects (I,D,E,F), then ten objects (II D,E,F); and that they would then learn to compare sets, first by one-to-one correspondence (VII B,C,D) and then by counting (VII E,F,G). The empirical analysis yielded three independent linear scales. Scale 1 includes all of the objectives for counting to five, in the predicted order, but also suggests that children learn rote counting to ten (II D) before they learn to count five objects. One objective for comparing by counting (VII E) falls into this scale. However, the objectives for counting objects to ten (II E and F) do not. Instead they appear in Scale 2 along with comparing by one-to-one correspondence (VII B and C) and the other comparing by counting objective (VII F). One objective (VII D) did not fall into either scale and appears by itself as Scale 3.

There are several difficulties in interpreting these results. Some difficulties derive from MSA's restriction to independent linear scales. For example, it is unlikely that counting objects to ten (II E and F) is truly independent of counting to five (I E and F). In MSA, however, the tests could not enter Scale 1 unless they also scaled with objective VII E. A possible hierarchy for these objectives is an upward branch in which counting to five leads both to counting to ten and to comparing sets: i.e.



However, using MSA, this hypothesis could have been tested only by running two separate analyses I D, II D, I E, I F, VII E, VII G; and I D, II D, I E, I F, II E, II F. Similarly, comparing via one-to-one correspondence may be prerequisite to comparing via counting, although not to simple counting. Here a downward branch can be proposed in which both one-to-one correspondence and counting are prerequisite to comparison by counting.



Again, however, this hierarchy is not directly testable under the assumptions of MSA.

Another source of difficulty in interpretation derives from the use of so many separate tests for closely related objectives. Possibly, by combining related behaviors we might produce more stable measures of the key classes of behavior and thus generate more easily interpretable scales. To explore this possibility, we next combined all tests of counting to five and gave a single pass or fail score for the set of tests. The same was done for the tests of counting to ten. Similarly, we computed a single score per subject for all tests covering the use of numerals to five and another for the numerals to ten. Finally, tests for comparing sets were combined to yield one score for the counting method and one score for the one-to-one correspondence method. These six summary scores were then analyzed using Multiple Scalogram Analysis. The results appear in Table 5.

In this analysis, all of the objectives involving counting fall into a single, quite easily interpreted scale. According to this scale, skill in counting objects is acquired before the numerals are learned (I before II, and III before IV), but both counting and numerals to five are learned before the child learns to count to ten. Comparison of

sets by counting is acquired only after basi counting and numeration are established. Comparison of sets by one-to-one correspondence (V) appears in this analysis as an independent class of behaviors, neither dependent upon nor prerequisite to counting and numeration skills. This finding seems reasonable with respect to simple counting and numeration skills (Objectives I – IV). However, it seems unlikely that the two comparison skills (Objectives V and VI) are completely unrelated to each other. In the MSA program,

TABLE 5

COMPARISON OF THE HYPOTHESIZED AND EMPIRICAL SCALES
BASIC NUMBER CONCEPT UNITS
(N = 37)

Hypothesized Scale		Empirical Scale	
		Scale I	Scale II
Objective I	(Counting objects 0–5)	I	V
Objective II	(Using numeral representation 0–5)	II	
Objective III	(Counting objects 6–10)	III	
Objective IV	(Using numeral representation 6–10)	IV	
Objective V	(Comparison of set size by one to one correspondence)	VI	
Objective VI	(Comparison of set size by counting)		
Reproducibility		.957	1.000

once Objective VI was shown to scale with Objectives I through IV it could not be considered for membership in a scale with Objective V. Although a separate program run for Objectives V and VI alone would have been technically possible, the assumptions of the Guttman scaling procedure make the testing of two-item scales a questionable procedure. Thus, there was no acceptable means, within the "scalogram" framework, of testing the hypothesis of a conjunctive branch in which counting and numeration to 10 (Objectives III and IV) and comparison of sets by one-to-one correspondence (Objective V) are prerequisite to comparison by counting (Objective VI).

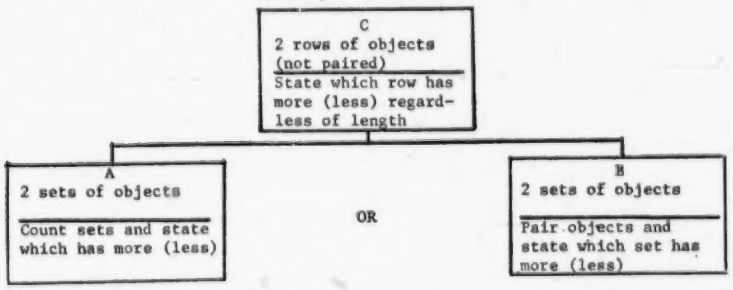
The repeated awkwardness of Guttman's scaling procedures in dealing with branching hierarchies led us to search for an alternative validation method whose assumptions would more closely match those of our hierarchical theory. Our requirements were the following:

1. Our hierarchies are generated one level at a time, by first identifying components of the terminal behavior, next identifying

prerequisites of these components, then prerequisites of the prerequisites, and so on in a succession of individual "analyses." This means that the critical relationships in a hierarchy are those between vertically adjacent items, (e.g., Figure 1, between F and H, E and F, C and H, E and G, etc.) rather than across an entire scale. Thus, it was appropriate to seek a method of validation that tested these adjacent relationships directly and did not immediately seek to construct multi-test scales or summary statistics covering an entire hierarchy.

2. The validation method should provide a means of testing several kinds of branches. These include (a) upward branches, in which a single objective is prerequisite to two or more higher level objectives (e.g., in Figure 1, E is prerequisite to both F and G); (b) downward conjunctive branches in which several objectives are jointly prerequisite to a single higher level one (e.g., in Figure 1, F and C must *both* be learned before H can be learned); (c) downward disjunctive branches in which *either* of several objectives is a prerequisite to a higher level one. Figure 3 shows a downward disjunctive branch. The hierarchy hypothesizes that in order to compare the number of objects in two rows (C) the child can either count the sets (A) *or* use a method of one-to-one correspondence (B). He need not, however, be able to perform both A and B.

FIGURE 3
A DISJUNCTIVE BRANCH


C
2 rows of objects
(not paired)
State which row has more (less) regardless of length
A
2 sets of objects
Count sets and state which has more (less)
OR
B
2 sets of objects
Pair objects and state which set has more (less)

3. The method selected should, ideally, permit a process of "search" among objectives for hierarchical relationships not previously hypothesized. These would in effect provide hypotheses for subsequent studies. While this is not a theoretical requirement, the

possibility of such searches would be a valuable tool during the early stages of research in a new area. This capability will of course require a computerized analysis capable of handling large quantities of data and of considering many alternative relationships.

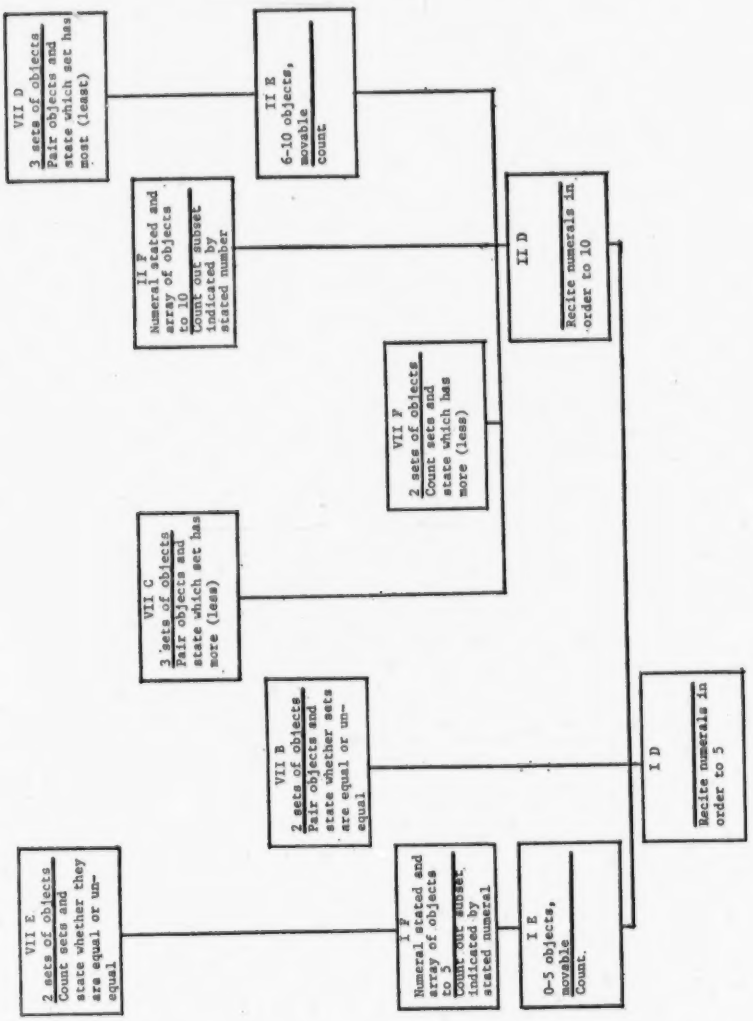
Other investigators have used procedures that met the first of these requirements. Gagné's various hierachy studies[10,14,13] used pass and fail contingencies for adjacent objectives in a hierarchy to compute a "proportion of positive transfer" statistic—essentially the inverse of the percentage of cases in which an individual passes a higher level test while failing the lower level "prerequisite." Walbesser's[34] proposed method for validating the AAAS science curriculum also uses pass-fail contingencies to test the "dependency" of each individual objective on its immediate prerequisite. Both Gagné and Walbesser directly test downward conjunctive hypotheses by combining data for two or more prerequisite tests and assigning a "pass" score only if all tests are passed. Upward branches are not tested directly, but are in effect implied when each of two higher-level objectives is shown to have the same lower-level objective as its prerequisite. However, neither Gagné nor Walbesser has discussed methods of testing downward disjunctive branches. Finally, neither of these methods is appropriate for empirical construction of hierarchies from test data, as opposed to validation of deductively analyzed hierarchies.

Dr. John Carroll, of ETS in Princeton has developed a hierarchy validation procedure that meets the requirements outlined in paragraphs 1 and 2, and which will also be, once a computer program is completed, quite economical to apply to large quantities of data, thus permitting empirical search for hierarchical relationships.[1] Carroll's method, like those of Gagné and Walbesser, begins with the construction of pass-fail contingency tables for all possible pairs of items in the hierarchy. Phi/Phimax* coefficients are then computed for each table. When the coefficient reaches an acceptable level, a hierarchical relationship between the two items is inferred, with the test showing the higher pass rate considered prerequisite to the one

* "Phi" is essentially an estimate of the correlation between two tests, each scored dichotomously. Phimax is an estimate of the highest-possible phi coefficient given the marginals of the contingency table. Since phimax would become larger as the pass or fail rate of either test became more extreme, the use of phimax in the denominator essentially controls against artificial inflation of the association due to extreme pass or fail rates.

FIGURE 4

HIERARCHY OF COUNTING AND SET COMPARISON SKILLS ACCORDING TO CARROLL'S ANALYSIS



with the lower pass rate. On the basis of these simple prerequisite relationships, it is possible to construct a hierarchy which can have both linear and branching sections.

Figure 4 shows a hierarchy derived from applying Carroll's program to the data analyzed in Table 4. The hierarchy contains both upward branches and downward conjunctive branches. Each of these types of branches can be logically derived from the simple prerequisite relationships.* Downward disjunctive branches, however, must be tested directly. The Carroll program will do this by combining two tests and giving them a pass score if either of the two tests was passed. Phi/Phimax coefficients will then be computed for these new scores. Since the computer program for disjunctive contingencies has not yet been completed, and hand calculation is extremely tedious, we have not yet applied this analysis to our data. However, we believe that the study of alternate routes to learning objectives—the essence of the disjunctive hypothesis—may be one important means of accounting for individual differences within a hierarchical framework.

The hierarchy in Figure 4 shows many branches, with very short linear paths. It is in some respects easier to interpret than the scales shown in Table 4. Essentially, the hierarchy breaks up Scale 1 of Table 4, showing rote counting to ten (II D) as not prerequisite to counting objects to five (I E and F), but as dependent upon rote counting to five (I D). This is precisely what would be expected from a behavioral and logical analysis of counting skills. On the other hand, the hierarchy also shows the five tests of Scale 2 as being unrelated to one another. This result is not so easy to interpret; behavioral analyses would have predicted that VII C would remain dependent upon VII B, and II F on II E. Further testing using new subject samples and, where necessary, revised tests, will be needed both to clarify the substantive issues raised here and to further explore the characteristics of Carroll's validation method.

In the research discussed up to this point, attention has focused exclusively on the possibility of predicting lower level behaviors from performance on higher level ones; no attempt has been made

* Direct testing of downward conjunctive branches is not *logically* necessary. If a test is independently dependent on each of two other tests, then it cannot logically be passed unless *each* of its prerequisites is passed. Nevertheless, Carroll is planning to include an empirical check on this deduction by combining two or more tests to yield a single pass or fail score and then computing phi/phimax coefficients for the combined scores.

in these studies to directly study the effects of learning lower level, presumably prerequisite, skills on the learning of higher level behaviors. To study these transfer effects, experiments involving instruction in the elements of the hierarchy are required. Such experiments by directly inducing acquisition of certain behaviors, permit more direct tests of transfer hypotheses.

Gagné[10] reported an exploratory study in which ability to perform a terminal task, given verbal directions only and no "practice," was measured before and after completion of a hierarchically arranged teaching program which stopped short of the terminal objective. This study in effect measured transfer to the terminal task from all of the subordinate learning sets combined. Other studies by Gagné,[14,13] as well as a more recent study by Ford and Meyer,[9] use a combination of instruction and scale analysis to test transfer among the subordinate sets themselves.

In each of these studies subjects worked through a teaching program designed to teach each of the behaviors in the hierarchy. Although the programs were designed to teach with a minimum of errors, demonstrated mastery of one unit was not required in order to move to the next unit. Thus it was possible to "complete" the program without mastering all of the behaviors taught. Upon completion of the program subjects were tested on mastery of each separate behavior in the hierarchy. The data were examined to determine the percentage of subjects able to perform each behavior who were not also able to perform the predicted prerequisites for that behavior—in effect for scaling "errors." A low rate of such errors indicated that mastery of the "prerequisite" was needed in order to profit from direct instruction in the higher-level objective and thus confirmed the hierarchical hypotheses.

A study by Merrill[24] introduced a mastery criterion into the teaching program itself as a means of testing the transfer characteristics of a hierarchy. Some subjects were given correction and review on successive tasks within a program until they reached a criterion of mastery; other subjects continued through the program regardless of mastery of the successive tasks. Merrill assumed, in accord with hierarchical theory, that mastery of lower level tasks would produce faster, more accurate learning and better retention of higher level tasks. He thus predicted that the correction and review group would go through the program more quickly and would perform better on immediate and delayed post-tests than the other group. These pre-

dictions were not borne out, and Merrill concluded that mastery of tasks lower in a hierarchy is not essential to learning a higher level task. It should be pointed out, however, that the hierarchy on which Merrill's teaching program was based had not been independently validated. Thus, Merrill's results may simply mean that the particular hierarchy studied is invalid rather than that hierarchically ordered sequences in general do not produce positive transfer.

All of the studies just described have attempted to study transfer properties of an entire hierarchy, and each has used a fairly extensive teaching program as its instructional vehicle. An alternative strategy is to study transfer relationships between adjacent pairs of behaviors in a hierarchy or among short sequences of behaviors. This strategy, while requiring many more separate studies than the total hierarchy approach, permits much tighter experimental design. In addition, as Gagné[12] has pointed out, it puts hierarchy research in contact with a past body of psychological research in transfer variables. A number of experimental designs for such small-scale transfer studies are possible.

One such design is to teach several behaviors in each of several different orders to different groups of subjects and to take repeated measurements of achievement of all behaviors during the course of instruction. Uprichard[33] used this approach in studying various sequences of instruction for the basic mathematical concepts of "greater than" (G) "less than" (L) and "equivalent to" (E). Six groups of nursery school children received small group instruction in these three concepts, each group learning the concepts in a different sequence. A test covering all three concepts was administered at the end of each week of instruction. When three out of the four subjects in a group reached criterion on the concept being taught, the entire group moved on to the next concept in its sequence. The week-by-week test scores on each concept for each of the groups provided the basic data in this study. Only the groups who were taught E first reached criterion on a concept in the first week of instruction. The groups beginning with G and L reached criterion on E in the third or fourth week of instruction without ever being taught the concept directly. The groups beginning with L learned only E in four weeks of instruction and had not learned L when the experiment ended. Thus, the data make it clear that E is the easiest to learn of the three concepts and L the hardest. The group taught in the order E–G–L was the first to reach criterion on all three concepts (in the

fourth week), thus suggesting that this is the optimal order for teaching the three concepts. However, the data is not absolutely clear in this respect, since the G–E–L group reached criterion on both G and E in the third week at a time when the E–G–L group had still acquired only E.

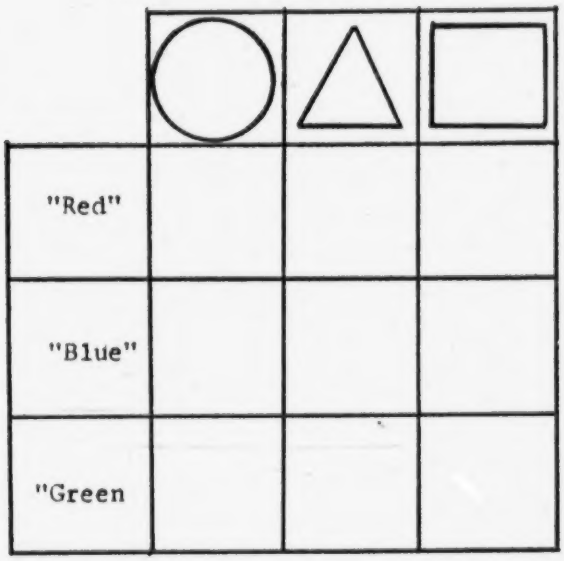
A more sensitive measure of learning is available when subjects are run individually; trials to criterion or error rates on each task in the learning situation itself can then be used as the dependent variable. Assume that two behaviors are taught in two orders, A–B and B–A, to two groups of subjects. According to hierarchical theory, if B is dependent on A then trials to criterion for task B in the order A–B should be significantly lower than for the same task in order B–A. An additional implication is that in order B–A, A should be "learned" virtually without error in the formal presentation, since the subject must somehow have learned A on his own in order to have acquired B. Finally, the total number of trials for tasks A and B combined should be lower in A–B than in B–A order, since the former would be a more efficient order in which to teach the set of tasks.

A recently completed experiment by Resnick, Siegel and Kresh[31] used this design in a study of double-classification skills in young children. Two tasks were used. Both required the child to correctly place objects in the cells of a matrix. In task A the defining attribute for each row and column was "given" to the child in the form of a filled "attribute" or "edge" cell. In task B, there were no attribute cells and the subject had to infer the defining attribute from filled interior cells in the matrix. A typical matrix for each task appears in Figure 5. We hypothesized that task B was dependent upon task A. In accord with the predictions just outlined, our results showed significantly more trials to criterion for task B when it came first than when it was preceded by task A. In addition, the predicted "immediate" learning of task A in second place did occur for subjects who had succeeded in learning B. However, the number of trials to criterion for the two tasks combined was not significantly different for the two orders.

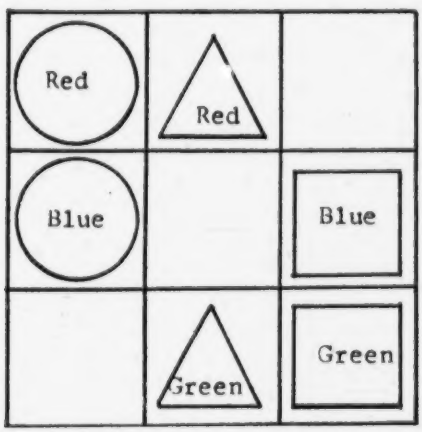
Members of our staff are now designing several other transfer experiments which will be run over the next several months. We view such studies as a means not only of ordering specific behaviors, but also of exploring the relations between hierarchical sequences and actual teaching procedure. For example, we intend to explore the conditions under which practice on a terminal behavior may be

FIGURE 5
SAMPLE MATRIX TASKS

A



B



more efficient than learning a hierarchical set of subordinate behaviors. We will also want to ask, as we have begun in the study just reported, what effect practice on the terminal behavior has on learning subordinate behaviors. Eventually, as the parameters of transfer in learning hierarchies become clearer, we hope it will be possible to define individual differences in learning as a function of the ways in which hierarchical structures are acquired. Some individuals, for example, may be able to skip over certain behaviors in a hierarchy while others may need explicit instruction at every step. Similarly, some may need extensive practice, to the point of "overlearning," before a newly learned behavior facilitates learning of a higher level objective, while others may show transfer effects from brief exposure.

With respect to applied work in curriculum design and evaluation, our work will continue to be concerned with defining and sharpening the role of hierarchical analysis, and in particular with determining the extent to which scalability of tests accurately predicts transfer relations among the behaviors. To explore this question, it will be necessary to conduct both psychometric studies, in which batteries of tests are administered and examined for hierarchical relationships, and experimental training studies, in which the behaviors in question are taught and transfer effects evaluated. By conducting both types of studies on each major hierarchy investigated, we expect to be able to examine empirically the extent to which scaling properties of hierarchies have direct implications for teaching sequences. We will also be able to explore the extent to which varying teaching sequences can produce differing scale structures. As these relationships become clearer, behavior analysis and learning hierarchies can be expected to become increasingly more valuable tools in educational research and development.

REFERENCES

1. Carroll, J. B., Personal communication, March, 1969.
2. Cox, R. C. and Graham, G. T. The development of a sequentially scaled achievement test. *Journal of Educational Measurement*, 1966, 3, 147–150.
3. Edwards, A. L. On Guttman's scale analysis. *Educational and Psychological Measurement Journal VIII*. 1948, 313–318.
4. Edwards, A. L. *Techniques of attitude scale construction*. New York: Appleton-Century-Crofts, 1957.
5. Elkind, D. Children's discovery of the conservation of mass, weight, and volume: Piaget replication study II. *Journal of Genetic Psychology*, 1961, 98, 219-227.
6. Ferguson, R. L. The development, implementation, and evaluation of a computer-assisted branched test for a program of individually prescribed instruction. Unpublished Doctoral dissertation, University of Pittsburgh, 1969.

7. Festinger, L. The treatment of quantitative data by "scale analysis." *Psychological Bulletin*, 1949, *44*, 146–161.
8. Flavell, J. *The developmental psychology of Jean Piaget*. New York: Van Nostrand, 1963.
9. Ford, J. D., Jr. and Meyer, J. K. A test of Gagne's hypothesis of hierarchical subtasks for developing learning program and an alternate proposal. Paper presented at Western Psychological meeting, April 28, 1966, at Long Beach, California.
10. Gagné, R. M. The acquisition of knowledge. *Psychological Review*, 1962, *69*, 355–365.
11. Gagné, R. M. Curriculum research in the promotion of learning. *Perspectives of Curriculum Evaluation*, 1967.
12. Gagné, R. M. Learning hierarchies. *Educational Psychologist*, 1968, *6*, (1).
13. Gagné, R. M. and Paradise, N. E. Abilities and learning sets in knowledge acquisition. *Psychological Monographs*, 1961, 75 (Whole No. 518).
14. Gagné, R. M., Mayor, J. R., Garstens, H. L. and Paradise, N. E. Factors in acquiring knowledge of mathematical task. *Psychological Monographs*, 1962, 76, (Whole No. 518).
15. Glaser, R. Instructional technology and the measurement of learning outcomes. *American Psychologist*, 1963, *18*, 519–521.
16. Green, B. F. A method of scalogram analysis using summary statistics. *Psychometricka*, 1956, *21*, 79–88.
17. Guttman, L. A basis for scaling quantitative data. *American Sociological Review IX*, 1944, 139–150.
18. Holland, J. G. Research on programming variables. *Teaching machines and programmed learning, II data and direction*. R. Glaser, Department of Audio Visual Instruction, National Education Association of the United States, 1965.
19. Kofsky, Ellen A scalogram study of classificatory development. *Child Development*, 1963, 191–204.
20. Kohlberg, L. Early education: A cognitive developmental view. *Child Development*, 1968, 39, 1013–1062.
21. Kropp, R. P., Stoker, H. S. and Bashaw, W. L. The construction and validation of tests of the cognitive processes as described in the *Taxonomy of Educational Objectives*, U.S. Office of Education, Cooperative Research Project 2117 U.S. Department of Health, Education and Welfare, Institute of Human Learning, Department of Educational Research and Testing, Florida State University, Tallahassee, Florida, February, 1966.
22. Lingoes, J. C. Multiple scalogram analysis: A set theoretic model for analyzing dichotomous items. *Educational and Psychological Measurement, XXIII*, (No. 3), 1963.
23. Loevinger, Jane A systematic approach to the construction and evaluation of tests of ability. *Psychological Monograph*, 1947, *61*, (Whole No. 4).
24. Merrill, D. Correction and review on successive parts in learning a hierarchical task. *Journal of Educational Psychology*, 1965, *14*, (5), 225–234.
25. Miller, R. B. Analysis and specification of behavior for training. Glaser, R. (ed.) *Training Research in Education*. New York: John Wiley, 1965.
26. Newell, A. On the analysis of human problem solving protocols. In J. C. Gardin and B. Jaulin (eds.) *Calcul et formalisation dans les sciences de l'homme*. Presses Universitaires de Frances, 1968, 146–185.
27. Peel, E. A. Experimental examination of some of Paiget's schemata concerning children's perception and thinking and discussion of their educational significance. *British Journal of Psychology*, 1959, *24*.
28. Piaget, J. *The origins of intelligence in children*. New York: International University Press, 1952.

29. Resnick, L. B. Design of an early learning curriculum. Working Paper #16, Learning Research and Development Center, University of Pittsburgh, 1968.
30. Resnick, L. B. Behavior analysis and the generation of learning hierarchies. In preparation.
31. Resnick, L. B., Siegel, A., and Kresh, E. The sequence of acquisition of matrix classification skills. In preparation.
32. Spiker, C. C. The concept of development: Relevant and irrelevant issues. In H. Stevenon (ed.) *Concept of development: A report of a conference commemorating the fortieth anniversary of the institute of child development, University of Minnesota.* Monographs of the Society for Research in Child Development, 1966, *31*, (No. 5) (Serial No. 107), 40–54.
33. Uprichard, A. E. An experimental study designed to determine the most efficient learning sequence of three set relations in the pre-school years. Unpublished Doctoral dissertation. University of Syracuse, 1969.
34. Walbesser, H. H. A hierarchically based test battery for assessing scientific inquiry. Paper presented at the American Educational Research Association annual meeting. 1968.
35. White, B. W. and Saltz, E. Measurement of reproducibility. *Psychological Bulletin*, 1957, *54*, (2), 81–99.
36. Wohlwill, D. F. A study of the development of number by scalogram analysis. *Journal of Genetic Psychology*, 1960, 97, 345–377.
37. Wang, Margaret C. (ed.) *Criterion-Referenced Tests* for the early learning curriculum of the primary education project, University of Pittsburgh, 1968.

Some Problems with Regard to Research and Development in Higher Education

LELAND L. MEDSKER

Research in education in general is big business and the segment of higher education shares fully in the enterprise. Total appropriations for "Research and Training" by the Congress alone increased tenfold to a total of over a hundred million dollars in the period 1957–1969. If grants by foundations and other donors were considered, the amount would be considerably greater. In 1957 relatively few major research projects pertaining to higher education were in progress. The *Inventory of Current Research on Higher Education—1968,* a project sponsored jointly by the Carnegie Commission on the Future of Higher Education and the Center for Research and Development in Higher Education, University of California, Berkeley, listed nearly 1,000 projects and had contacts with more than 2,000 researchers.[1] Since the *Inventory* listed only research underway and not reported, it does not portray the total volume of inquiry and findings that annually bear on education beyond the secondary school.

The rapid increase in research in higher education is reflected in many forms and types of activities. There is research within colleges about themselves—in other words, institutional research—which has become nearly universal in practice. Then there are the hundreds of individual researchers engaged in inquiry on various facets of education at the college level. Recent years have witnessed the development of organized research units, usually but not always based in universities, which systematically attack major problems in the field. More will be said later about the research and development center concept which was initiated by the United States Office of Education in 1964 and which developed from a concern that the fragmentary

nature of most educational research did not lead to sufficiently cumulative findings for influencing change. The alternative was to establish a limited number of centers in which programmatic research on crucial problem areas would be conducted. To date nine such centers have been funded by OE, although the Berkeley unit is the only one focused entirely on higher education. Inherent in the center idea is the concern for translation of research findings into practice, hence the emphasis on *D* or development, about which more is said later.

The prevalence of research in higher education is further reflected by the extent to which the Educational Research Information Center (ERIC) system has expanded as a means of storing and disseminating findings. An ERIC Center for higher education was established at George Washington University in 1968. It is supplemented by other Centers—such as the one at UCLA on junior college research data, and one at the University of Michigan concerned with research findings on student personnel.

Thus at the end of the decade of the 1960's there is evidence of widespread research and development efforts in higher education. The need for research in the field increases annually and the problems connected with it become ever more complex. The remainder of this paper will attempt to deal with some of these problems and their implications.

As background for the problems let us turn first to the changing scene in higher education—a scene which naturally reflects a total society that is itself far different today than it was only a few years ago and whose future is destined to still further radical change. Higher education is a vast enterprise. It requires the support of 2 per cent of the gross national product and involves about 4 per cent of the population, including more than a third of the college age group. With the goal of near universal education through at least the junior college years the growth spiral continues upward. And so does the cost spiral! The financial plight of the private colleges and the manifest resistance on the part of the public for the support of public institutions raises serious questions as to how the enterprise is to be sustained. But other issues loom with equal intensity, some alarmingly, others auspiciously. They are familiar to all: Student rejection of the status quo; confrontation; violence; questions concerning the education of various ethnic groups; cleavages between faculty, administration, students, and governing boards, and disenchantment on the part of the public. Efforts to meet these problems are many

and varied. They include curricular innovations, new configurations of governance, greater emphasis on institutional and state planning, and often sheer compromise as a means of maintaining peace on the campus.

This is the context in which research in higher education must take place and it makes certain problems immediately apparent. One such problem is the increasing variability of the research variables. Nothing is static. The nature of student input changes constantly. The purposes and goals of higher education seem to shift from one period to another and are perceived differently by various participants. Changes over the period of a longitudinal study which covers the college years may invalidate any controls of variables or may significantly alter the nature of the study. A special problem is encountered in those projects which attempt to assess the impact of college on a graduate's performance in life activities by the fact that societal situations a few years following college may be entirely different from those for which the college experience was designed. In fact, a question could be raised as to whether some of the earlier studies such as those conducted at Vassar would be valid in the present climate of change. The matters that are of primary concern about students today are different from those which engaged us yesterday. Until recently a study of college students might have included an assessment of their attitudes about sex and liquor whereas today these issues may seem pallid when compared with drugs and violence. Likewise, any earlier study of the decision-making process in colleges and universities that would have wrestled with the problem of authority between faculty and administration must now be concerned with student involvement in a variety of forms.

A second problem is what one might term the shifting relevance and significance of issues for research. During the last few years—indeed the last few months—developments in the nation's colleges and universities have raised serious problems which were not prominent theretofore. Some of them are so grave as to question the very survival of our social institutions and most of them suggest new priorities for investigation. Take, for example, the question of how different ethnic groups are to be served. Just as colleges seemingly were finally about to take steps toward integration, it now develops that separate departments, if indeed not separate institutions, are to emerge as the way of serving these groups. This comment implies no value judgment as to how what is probably the most important, diffi-

cult, and belated task in higher education is to be performed but it does suggest that new ways and forms of serving these students should be a research topic of high order. A similar question relates to admission policies pertaining to students who by the usual criteria do not meet the entrance requirements of selective institutions. What is the impact on both the students and the institution when the usual standards are waived? What new criteria are needed for evaluating prospective students from culturally different backgrounds? Consider other emerging issues such as the impact of confrontation and compromise on the institution or the matter of student involvement in decision-making, whether by forceful demand or peaceful assimilation, and at once it is clear that the results need assessment by means other than mere guesses. In the same vein one can consider such questions as the following: the effect of federal financial aid to students (effects on both the students and the institutions), problems associated with what seems destined to be a greatly expanded system of non-baccalaureate institutions (new types of vocational schools as well as community colleges), emerging governance and planning configurations, innovative efforts to reorganize the undergraduate curriculum, and new concepts of graduate education. The list is grossly incomplete, but it takes little imagination to realize that unless such new issues become the concern of both researchers and practitioners, Rome is in danger of burning while many people fiddle.

It is evident that an increasing amount of research must be of a kind that is useful to decision makers, planners at local and state levels, and legislative bodies, including Congress. The ivory tower research concept is due to decline, if it ever existed. The day for identifying and attacking the crucial problems and of leading toward solutions is at hand. This, however, is the difficult way to plan research.

Still another problem that complicates research today is the matter of constraints which are imposed by the public in an effort to avoid the invasion of privacy. For some time, any agency conducting research under a federal grant has had to submit any instruments to be used in gathering data to the Office of Education for approval. While our own experience with OE has been positive in that its staff has responded quickly to our many requests for reviews and in general has been liberal in its approval of controversial items, it is nevertheless a process which requires a great amount of time and planning. But the same caution is now extended into other situations with

even greater constraints. For example, in California it is now illegal to administer certain types of instruments to high school students without prior parental consent. Such a restriction has imposed great difficulty for the Center's SCOPE* project under the direction of Dale Tillery in which he is to follow up some 9,800 high school students in the state this spring. True, the requirement at the moment applies only to students below the college level, but two matters are of concern. In the first place, one never knows when similar legal restrictions may be extended to the college level. And secondly, many studies in higher education involve contacts with secondary school students and thus the restrictive policies apply automatically.

Still another restriction arises out of those policies and laws which prohibit the identification of students as members of ethnic groups. While one can appreciate the rational behind such restrictions, it could be argued that they often tend to preclude the very research from which the groups are most likely to profit.

It is true, of course, that some of these are fundamental problems characteristic of all educational research and that both researchers and people in the field have a responsibility for seeking a balance between the right of privacy and the advancement of social research. As Dr. Tillery said in 1966[2] at a symposium of the National Council on Measurement in Education in New York on this subject:

> In summing up, scientists should have the right to study human phenomena but also the responsibility to seek the cooperation of individuals and institutions in a manner which clearly respects the right of privacy and the protection of anonymity. This forces the investigator into very careful plans for seeking counsel and understanding of individuals and groups associated with his enterprise. It means the willingness and the ability to communicate the importance of the research and the basic rationale for the methods and techniques being used. If we are not willing to involve practitioners in our work, particularly those who carry the weight of responsibility for decision-making in a time of great social stress, we may be forced to restrict the kinds of investigations which we may conduct.

The problems discussed so far tend to fall into the category of externally derived constraints. They are imposed on the researcher by the very nature of the current environment and he has little choice other than to cope with them. There is another category which tends to stem from within the research world—though naturally this too is

* School to College: Opportunities for Postsecondary Education

affected in part by developments outside. In this group can be included the complex factors that arise out of the many and varied research and data processing methodologies and the current emphasis on the interdisciplinary approach. The range of acceptable research techniques is far greater today than it was a mere few years ago. Moreover, the age of the computer and its affiliates now make it possible to initiate projects of a magnitude that could not have been conceived a decade ago. Under these circumstances hard decisions have to be made as to the most appropriate and feasible project dimensions as well as the techniques to be used. The decision as to whether to use the micro or the macro approach to a problem is no longer based on the question of whether the latter is possible so much as it is on a consideration of which will be the more appropriate.

The idea of interdisciplinary research is now popular and for good reason since it is agreed that the background and approach of researchers in various fields, particularly in the behavioral sciences, should be brought to bear on problems in higher education. But the task of organizing a team of researchers representing several disciplines is easier said than done. Often the individuals either do not wish to cooperate on the same project or they do not have the temperament to do so. A representative from a discipline may engage in research on a given educational problem and may even confer from time to time with his peers in other disciplines, but this alone is not interdisciplinary research. Naturally, there are many examples of teams from various fields which are successfully mounted, but they are the exception and the process is difficult.

If the foregoing identification of certain difficult problems of research in higher education can be accepted, the problems should now be viewed for their implications to individuals and units engaged in research. Let us first examine their relevance to individuals who are attached to colleges and universities and who are involved in conducting research either by themselves or as members of an institutional research unit. A number of considerations come to mind. In the first place, everyone has the problem of determining the relative significance of potential research projects and of attacking those which seem to be most in need of study in light of today's societal perplexities. Anyone can keep busy doing research of interest to him and undoubtedly to others, but since there is neither enough money nor talent to attack all the most serious problems, some value judgments have to be made concerning those to which the resources

should be allocated. It would appear that too many bureaus of institutional research engage in various types of "head counting" which yield little more than "nice-to-know" information and really do not make much impact on the institution.

A second guideline is that projects undertaken must be manageable. With present research technology there is the temptation for researchers to bite off more than they can chew. This is true of graduate students as they undertake dissertation studies and it is a disease to which we are all susceptible. Often research results will be more meaningful if they are pinpointed and if the findings are reported quickly and without undue complexity.

Another possibility is for representatives of groups of institutions to organize themselves into consortia arrangements so that they can assemble comparable data from across the institutions represented and thus have broader bases for comparisons and generalizations. In a sense this idea is in opposition to the preceding one that projects should often be small. However, in view of the data processing capabilities available today, it is possible to have both large-scale and small-scale projects with each filling a particular need. Another possibility for participation in large-scale projects which still involve one's own institution is through cooperation with organized research agencies on specific projects thereby gaining the benefit of comparable data from many institutions and the assistance of the staff of the large research agency which is often geared to the macro approach.

Perhaps the most important implication of all is that data coming with comparative ease from many projects need to be carefully interpreted for their implications for individual institutions. The sheer quantity of data sometimes leads to hasty generalizations or sometimes to none at all. The opportunity for individual researchers or for those in bureaus of institutional research to compare their own findings with those of others and to postulate further implications is great and the individual researcher who fails to take the comparative stance is forfeiting an opportunity to make his own research of greater significance.

Naturally, many implications can be drawn for large-scale organized units based in universities or elsewhere. To some extent, the implications for these units are similar to those pertaining to individuals or institutional research bureaus, but in other ways they differ considerably simply because of the nature and size of most organized units. These centers, for example, have the same problem of deter-

mining the relative significance of issues in need of research and thus of setting priorities for themselves in terms of their program. As Norman Boyan[3] said in a major address at the American Educational Research Association conference in Los Angeles:

> What we know now as never before is that we must also make sense out of the following questions: What are your *substantive* priorities? What significant problems are you trying to solve? How do you propose to allocate your resources to solve these problems? What timeframe is necessary for solving these problems? What evidence will you accept that you are moving toward solution of these problems?

On the other hand, because of their overall capacity the centers inherit the even more fundamental problem of determining the balance between basic research and other types of inquiry. It would be entirely appropriate for a unit to undertake an exceedingly basic study on the learning process at the college level, but whether to do so may depend upon the press for solutions to other current problems in higher education, solutions that suggest a greater emphasis on applied or policy-oriented research. The criteria for making such a determination will vary with the objectives of the research center and the sources of its funding.

Obviously, it would be lamentable if in their zeal to deal only with the problems of the day, research centers were to omit entirely any consideration of contributing to knowledge through research on some of the fundamental problems of educating people. On the other hand, if the crucial problems confronting the colleges today are to be solved in the light of rationality there should certainly be some input from research and there is a good question concerning who will make this input if organized centers do not. It is true, of course, that in some areas basic research and other types of inquiry are not mutually exclusive and thus basic research does not preclude investigation leading to policy determination.

The organized center also has the opportunity and probably the responsibility to conduct its research on a programmatic basis so that one step tends to follow another and also so that the total effort in researching a problem area is coordinated, even if several projects are involved.

The organized centers also face all the problems inherent in the interdisciplinary approach. Because of their size and the fact that they often reside within universities, the use of an interdisciplinary

team ostensibly is relatively easy for them; on the other hand, their utilization of this process is subject to all the problems and limitations referred to earlier.

Perhaps the most crucial problem of all in an organized center, and particularly one funded as a research and development agency, is the relationship between research and development. As a matter of fact, there is a prior question of just how development should be defined. Generally speaking, it is presumed to be those efforts which effect change in education, particularly by the use of research findings. While research and evaluation may stem from educational practice in the field, the more general notion is that research findings need further experimentation and that change should be expected to follow from research. In any event, the efforts of an R and D center are presumably directed toward decision makers and this poses two problems. One is that of determining how much of a center's time and effort should be devoted to development and the other is how a center can best reach and influence decision makers.

An agency devoted to the study of higher education has a particular problem with respect to the latter question in that the decision makers in this segment are exceedingly diverse and dispersed. They include faculty, students, administrators, members of governing boards, statewide coordination agencies, and legislative bodies. Nonetheless, with the mounting problems in higher education an organized research unit in this field, whether or not expected to do so by its funding agency, is obligated to be concerned with the process of effecting change despite all the problems involved.

During the last few years some good questions have been raised about development and many researchers have been concerned with the expectation in certain circles that research will immediately produce the means by which education can be revolutionized. Other researchers have feared that an overemphasis on development would militate against their research productivity. In his AERA-PDK[4] Award Lecture at the Annual meeting of the American Educational Research Association in 1967 in New York, T. R. McConnell made some sound remarks about this problem:

> I should say that an interest in development does not necessarily endanger educational research, either basic or applied. It is pressure for a quick pay-off—for an educational cookbook, some more hardware, for a new and magic educational nostrum—that threatens both significant research and sound development. The notion in some quar-

> ters that it is only a jump from either basic or applied research to improved educational practice is much over-simplified. Experience in other fields has shown that many processes intervene between research and production. It also has been demonstrated that evaluation must accompany development. The transition from research to practice is not one leap. It is a process, a flow, from basic through applied investigation, to invention and development, to innovation in practice or production, and, finally, to evaluation. Without evaluation, development may easily become quackery.

McConnell's statement helps to clear the air about the relationship between R & D, but most organized research units must still struggle with the daily problem of how and what to do as a means of bringing research findings to bear on practice.

We might refer briefly to the Berkeley R & D Center as an example of a university-based organized research unit that is concerned with most of the problems outlined above. The Center was established in 1956 and for nine years was known as the Center for the Study of Higher Education. During this time it operated on a reasonably small-scale basis delving primarily into the problems of student development, but also to some extent into institutional analysis, statewide coordination, and related matters. In 1964, when the federal government announced its intention to establish a number of R & D centers, the Berkeley unit was invited to submit an application for funds to become such a center which it did in September, 1965. Since then its two major research foci have been (1) continuation of the earlier interest on the impact of college on student development and (2) college organization and administration with considerable emphasis on planning. Its research program has been accompanied by a strong emphasis on development and dissemination. The Center has tried to be both programmatic and interdisciplinary in its approach to research, but in doing so has encountered the difficulties enumerated earlier. It has now almost completed an extensive examination of its program and a delineation of its role and scope for the next few years. In this process, it has attempted to sharpen the focus of both its research and its development program. It looks now as if its primary focus will be on how best to extend the learning environment and that the approach for this effort will be through the examination of new types of educational programs and of emerging governance configurations.

In the final analysis, of course, it is the users of research who are the most important parties of the enterprise. You, who are in the

field, have a heavy responsibility to help make research relevant and meaningful. While not everyone in this audience is engaged directly in higher education, many are and many more are by reason of being in secondary schools indirectly concerned. There are various ways by which your role can be enhanced. Let me mention three. First, you can help identify the problems which, in your judgment, are crucial and which need the input from research to help solve them. Once you identify the problems they can be communicated to those research agencies, either within your institution or elsewhere, that seem to have the greatest potential interest and capability to attack the problems. Second, you may stand ready to cooperate with other agencies in major research activities. After all, investigations cannot be done in a vacuum, but instead must be carried on in the higher education community itself. Third, you may cooperate in development activities in which further exploration, experimentation, or utilization seem necessary to validate research findings. As we move from an era in which research findings were relatively passive to one in which we believe they must be active, the opportunity for engaging in development activities will be far greater than it has been in the past. The future, then, should bring an expanded opportunity for researchers and practitioners to work as a team.

REFERENCES

1. Heckman, Dale and Martin, Warren, *Inventory of Current Research on Higher Education – 1968*. New York: McGraw-Hill, 1968.
2. Tillery, Dale, *Seeking a Balance Between the Right of Privacy and the Advancement of Social Research*. Journal of Educational Measurement, Vol. IV, No. 1, Spring, 1967, Supplement, p. 11.
3. Boyan, Norman, Associate Commissioner for Research, U.S. Office of Education, *Educational Research and Educational Policy*, Address at the American Education Research Association, Feb. 6, 1969, Los Angeles, California.
4. McConnell, T. R., Center for Research in Higher Education, University of California, Berkeley, AERA-PDK Award Lecture, Annual Meeting American Educational Research Association, New York, February 17, 1967.

Educational Research, Educational Development and Evaluation Studies

JOHN K. HEMPHILL

No one who is in touch with the problems of our troubled society can fail to be aware of the criticism that is directed toward our educational system. The validity of any specific criticism may be debated, but there are few who can defend the system as it is now functioning. Although we have many reasons to be proud of our past accomplishments in education, our future aspirations seem unlikely to be attained.

Educational research, educational development and evaluation studies are the major tools by which we can improve education. Unfortunately, confusion exists as to what can best be accomplished by each of these tools. The thesis of this paper is that the functions of educational research are distinctly different from the functions of educational development and, further, that evaluation studies in education have yet a third function. Moreover, each of these tools has distinct characteristics which need to be made very explicit. A clear notion of the characteristics and of the functions of educational research, development and evaluation will make it possible for us to move ahead toward the achievement of needed educational reform.

ACTION

If one is to perform a purposeful act, he must do three things and do them in sequence. First, he must formulate an intention, i.e., he must visualize a state of affairs which differs from the state of affairs he believes now exists, and commit himself to bringing about the new situation he visualizes. Second, he must operate upon his environment in a way which he calculates will produce the state of affairs

he desires. Third, after he has performed the operation he must compare the state of affairs that he has caused to exist with the state of affairs that he had intended to achieve. He then notes whatever discrepancies remain. These discrepancies serve as feedback to influence his next intentions and, thus, guide his further action toward his desired end. It does not matter whether a purposeful activity is a simple one, such as finding a pencil upon a cluttered desk or a very large and complex one, such as leading a Christian life. In either case, the same three basic steps must be taken in sequence, an intention must be formulated, an operation or set of operations must be performed and a comparison made at the end to evaluate the results.

The second or operational phase of a complex act is often made up of nested sets of subacts. Thus, in carrying out an intention of leading a Christian life, an individual, among other subacts, may join a church and, within this subact and as a further subdivision of this large subact, attend services each Sunday.

The three phases of purposeful action—intention, operation and comparison—correspond closely to the basic functions of educational research, educational development and evaluation studies. Educational research can contribute to educational reform by providing new knowledge to be mixed with experience (old knowledge) in shaping our intentions or (in more usual language) contribute to the setting of objectives or goals. Orderly change or reform is not possible if one cannot visualize a better state of affairs than the one which presently exists. New knowledge generated by successful research can enlarge the number and the attractiveness of alternatives which one may consider.

Educational development contributes to educational reform by providing new and more powerful ways of operating upon the educational environment. It creates new educational products and/or new human capabilities which, when properly applied can make significant changes in educational practice.

Educational evaluation produces evidence that can sharpen value judgments about the present state of affairs in education. Evaluation, in essence, is simply comparing two states of affairs, one of which is considered to be more desirable than the other. Evaluation always implies a value judgment.

In a general way, it is suggested that the three major tools available for accomplishing educational reform correspond one-for-one

with the three basic steps of purposeful action. Educational research (or considering the findings of research) is to be employed when one is in the process of formulating his intentions, that is, setting his objectives. Educational development provides a manner for operating systematically upon these intentions. Educational evaluation provides a way of assembling evidence from which one may judge how well the state of affairs he intended to bring about has been achieved.

Let us now examine certain distinctive characteristics of educational research, educational development and evaluation studies.

EDUCATIONAL RESEARCH vs. EDUCATIONAL DEVELOPMENT

Research and development in education are almost always confounded in the thinking of most persons. R & D has become fused and often is regarded as a unitary process. Yet, there are characteristics of educational development, both in its purposes and its methods, that clearly differentiate it from educational research.

Educational development is the systematic process of creating new alternatives that contribute to the improvement of educational practice. Educational research is a scientifically disciplined process of creating new knowledge relevant to education. The findings of research in education seldom can be used to improve education without doing a considerable amount of additional work. Research outcomes are most often reported in technical terms without reference to possible practical application. New knowledge usually becomes useful only after much transformation, adaptation, and mixing with other knowledge which has been gained from experience.

It seems unnecessary to discuss here the process and methods of educational research. The canons of scientific research in general have been made explicit to the point of common knowledge. These principles and methods which apply to educational research as well as to any other research are taught as part of most graduate curricula. Educational development, however, is not well understood.

Educational development can take very different forms since the activities that constitute development are quite varied. No one process has been identified that provides a "blueprint" for educational development, perhaps because at this time too little experience has accumulated to form a substantial base from which to judge or evaluate alternate processes. However, two trends that suggest the major emerging variants can be seen among the activities of those engaged in educational development. These are: The *product development* process and the *change support* process.

THE PRODUCT DEVELOPMENT PROCESS

The product development process seeks to bring about improvement in educational practice by creating materials, procedures, or devices which, when used as directed, are known to yield desirable and specified outcomes. The emphasis is upon creating a tested and proven "package" with appropriate supporting materials such as manuals of instruction, operator training material, teacher guides, et cetera. Thus, the outcomes of a product development process can be described as "packages of things" that have physical identity.

A basic assumption of the product development process is that school personnel will be sufficiently motivated to seek and utilize the new and possibly better materials or procedures. A major block to be overcome in improving educational practice through product development is the unavailability of tested and proved educational materials. It is assumed that better materials need only be made available in order that improvements in educational practice will occur. This assumption is shared by the old adage, "Build a better mousetrap and the world will beat a path to your door."

It is important to describe in some detail what product development in education means in terms of the tasks involved. Although the exact nature of these tasks and their sequence of performance will vary from product to product, at least a general pattern of activities can be observed.

The first step in such a process is the judicious selection of the product to be produced. This begins with the awareness of a need or problem for which the product might provide a full or partial solution and involves a very broad specification of the product's characteristics conceived in terms of objectives, costs, feasibility, etc.

The second general step in the product development process is to carefully review the state of the art and knowledge from which the product is to be developed. This includes scrutinizing research literature in all revelant areas, assembling valid practical experience, and estimating the costs and difficulties encountered in bringing together the elements essential to the development.

The third step is invention and design. This entails elaborating the product's specifications and fixing upon one or a very few alternative "models" of the product to be created.

The fourth step in the development process is to prepare a preliminary version—a "mockup" or "prototype"—of the product and to test or examine its performance. This version will only be partially

adequate, but will provide information critical to succeeding steps.

The fifth step is to analyze the preliminary test data, applying direct attention to their implications for redesign of the product.

The sixth step is to assemble a revised version of the product, which incorporates the experience from the earlier version, again subjecting this revised "model" to a performance examination. Steps five and six may be repeated any number of cycles before moving to step seven, depending upon how successful the design-test-feedback-redesign operation has been. Once, however, a model of the product is produced which appears to perform to specifications, work proceeds to field testing.

Step seven, field testing, is to design and conduct a rigorous test of the product in a situation which duplicates most of the known relevant characteristics of the operating environment. Specific data are gathered about the performance of the product within differing general environments that will yield the "limits" within which the product may be expected to perform.

The final step is that of operational testing. This differs from field testing, in that the group responsible for the development work retires from direct involvement in this further testing of the product. This step establishes the feasibility of releasing the product for normal operational use without constant supervision by its originator. Only after the last hurdle of field testing is the product judged to be ready for dissemination.

THE CHANGE SUPPORT PROCESS

Educational development that is conducted following the *change support* process directly addresses changing the practice of education. It emphasizes intervention in the behavior of educators. In general, material things are regarded as incidental or clearly subordinate to improved attitudes, skills, motives and abilities of people. The behavior to be improved includes group or organizational interaction of people as well as that of the individual educator. The focus of efforts is not limited to individual remediation, but may also include rearrangement of relationships among groups, as these in turn affect the behavior of individuals within them. A basic assumption of the change support process in development is that educational practice is improved by direct intervention in what educators do.

The process involved in educational development through change support emphasizes flexibility. Each human situation is different

from all others, and each must be met differently. No prescription can be written which will be effective in all or a majority of situations. As a direct consequence, the persons engaged in change support seldom make explicit in advance just what steps they will take toward their objectives. From this point of view, development is a continuing process, never to be completed, since improvement can never be said to reach a point where further improvement is not possible. Objectives are regarded only as temporary states in a continuously changing set of human relationships.

The activities of an educational developer guided by this process are also characterized by flexibility. One role that has been described explicitly is that of change agent. Such an agent attempts to stimulate interest in changing present practices, provides information about what is possible, and encourages those who are attempting to change or to make changes in others. In some sense he functions as a catalyst in a larger process. Another role for the developer using this approach is that of a coordinator. He strives to bring together persons or agencies where improvements in education might result from increased communication, or where the effectiveness of activities that are being performed relatively independently could be increased if they were done in concert. A coordinator's role may involve him in negotiation and politics—especially professional politics—and may require the ability to manage the use of power. Still another role is that of trainer. In this role, the developer acts as a super-teacher of school personnel, but not of school students. There are a number of techniques available to the trainer that can be described as specific entities. Among these are role-playing, sensitivity training, T-group techniques, psychodrama, etc.

DEVELOPMENT STRATEGY

It is not possible at this time to prescribe a best strategy for educational development because of our very limited experience with it. Most of the strategies being employed today appear to be mixed strategies with different emphasis upon one or the other of the two approaches described in this paper. The major factor that has influenced the strategy adopted by those now engaged in development work is the background and experience of the individual developers. Opinions and beliefs about how education can be improved far outweigh solid evidence based on evaluated experience. Most of the persons now engaged in directing educational activities have en-

tered their new work from a wide range of previous professional occupations and have been trained in a variety of academic disciplines. Generally, they have earned graduate degrees in education, psychology, or sociology. Many have backgrounds that include classroom teaching and school administration. Because of their widely differing backgrounds, it is readily understandable that the differing assumptions of the two approaches (*product development* and *change support*) have different appeal.

EVALUATION STUDIES

Evaluation studies are frequently confused with research. It is clear that many activities are shared by research on one hand and evaluation studies on the other, but one cannot be considered a simple subset of the other. Less confusion exists, however, between evaluation studies and educational development. Evaluation for the purpose of providing feedback is a subactivity in the development process. In terms of the analysis of purposeful acts described earlier, evaluation done as a part of development is simply a subset within the larger operation.

Let us return to the task of making explicit the relationship between research on one hand and evaluation studies on the other.

Evaluation studies imply comparison and decision about alternatives; by undertaking an evaluation study, one at once addresses himself to questions of value and utility. It may be objected, however, that this is a too idealistic view of the purpose of evaluation studies. In fact, the great majority of evaluation studies in education may not be concerned with the alternatives *per se*, but instead ask the simple question, "Does treatment X work?" At best, there may be an implicit assumption that, "if X does not work, we will have to try something else," but this is as far as thinking about alternatives may go. Nevertheless, and regardless of the lack of precision in thinking, providing information for choice among alternatives remain the basic purpose of evaluation studies.

The implications of primacy of utility in evaluation studies and the relative unimportance of such a consideration in research are profound. Although there are differences in points of view among behavioral scientists, an "ideal" research study is characterized by most, if not all, of the following*

* I am indebted to Richard Watkins, Program Coordinator, Far West Regional Laboratory for Educational Research and Development, Berkeley, California, for much of this material.

1. Problem selection and definition is the responsibility of the individual doing the research.
2. Tentative answers (hypotheses) to the problem may be available by deduction from theories or by induction from an organized study of knowledge.
3. Value judgments by the research are limited to those implicit in the selection of the problem.
4. Given the statement of the problem and the hypothesis, the research can be replicated.
5. The data to be collected are determined largely by the problem and the hypothesis.
6. Relevant variables can be controlled or manipulated, and systematic effects of other variables can be eliminated by randomization.

Almost the reverse of all these six statements characterize the evaluation story;

1. The problem is almost completely determined by the situation in which the study is conducted. Many people may be involved in its definition and because of its complexity, the problem initially is difficult to define.
2. Precise hypotheses usually cannot be generated. There are many gaps where the absence of verified knowledge must be filled with judgment and experience.
3. Value judgments are made explicit in the selection and the definition of the problem as well as in the development and implementation of the procedures of the study.
4. The study is unique to a situation and seldom can be replicated even approximately.
5. The data to be collected are heavily influenced, if not determined, by feasibility. Choices, where possible, reflect value judgments of decision makers or those who set policy. Gaps exist between data that are feasible to collect and the data that would be most useful to the decision maker.
6. Only superficial control of a multitude of variables important to interpretation of results is possible. Randomization to eliminate the systematic effects of these variables is extremely difficult or impractical to accomplish.

Evaluation studies are not just poorly performed or less rigorous research studies. In fact, they can and should be done with as much rigor and imagination as the best of research. However, they differ in that they are undertaken in response to a need to know the usefulness of some combination of old and new knowledge which has re-

sulted in the invention of an alternative to existing modes of action. Is a new method of training teachers an improvement over a presently used method? Is a specific Head Start program effective in preparing disadvantaged youngsters to enter school?

If we accept the proposition that the basic reason for undertaking an evaluation study is to develop information that will assist a decision maker in choosing rationally among the alternative courses of action, then, an evaluation study is to be viewed from a perspective quite different than that from which one might view a research study. The highly regarded research act, "refuting a null hypothesis," carries little or no useful meaning, since for a decision maker to know that he cannot reasonably consider some situation or condition which is not stated in the hypothesis provides little guidance for the choices he must make. Confidence in a conclusion, as represented by the research convention implied by the general acceptance of the ".05 or .01 probability level" as the criterion for "belief" of a research finding, is a luxury a decision maker seldom can afford. Rather more frequently he faces situations where any information more dependable than that provided by a "flip of the coin" is desperately needed. The concept of "sampling" a domain of problems, of which the unique problem the decision maker faces in making a particular choice is only one case, is simply not applicable in the decision situation, but is the foundation of research design.

Statistical decision theory provides a proper framework for understanding evaluation studies. Within this framework an evaluation study becomes a process of acquiring further information, or new information, that can be used by the decision maker. The decision maker's probability estimates of the consequences of a contemplated act can be modified as a direct result of the outcomes of the contemplated evaluation study. His expectation about the outcomes of the contemplated evaluation study also have an estimable probability. This fact makes it possible for decision makers to step back a step and make a reasonable decision about whether an evaluation study would likely be worth what it costs. Thus, the expected value of carrying out an evaluation study is determined by the same criterion that is used to judge the consequences of an action. This criterion is not the criterion of the research worker who finds his "payoff" in the creation of "new knowledge," but is the "payoff" of the consequences of action taken by a decision maker.

The major differences between evaluation studies and research

studies is not the subject of interest or the method of inquiry of the researcher and evaluator. It is to be found in the manner in which the outcomes of the two are used and regarded.

SUMMARY

In summary, it has been suggested that research, development, and evaluation are the tools we must use if we wish to reform and renew educational practice. These tools have an analogous relationship to the three steps or phases of any purposeful action: the formulation of an intention, the operation upon that intention, and the comparison of the intended states of affairs with a realized state of affairs. Each of these tools has unique characteristics which fit them to the different functions that must be performed if education is to be improved.

If we are to make orderly and rapid progress in improving our education system, we who assume some responsibility for the task are obliged to know our tools and use them with insight and skill.

The Challenge of Multi-Agency Involvement in Development

RAY JONGEWARD

This presentation will employ a five-step planning model as a framework for the following data and information. It is a simplistic model and consists of:

Step 1. Who hurts?
Step 2. Why do they hurt?
Step 3. Who has the aspirin?
Step 4. Why isn't the aspirin as big as the headache?
Step 5. How can we make a better aspirin?

First, some background on the Small Schools Program of Northwest Regional Educational Laboratory. The 27 per cent of the land area of the U.S.A. which is composed of Alaska, Montana, Idaho, Oregon and Washington is 80 per cent rural. With the exceptions of urban centers, especially the "strip city" called the Puget-Willamette trough which extends from Everett, Washington, to Eugene, Oregon, much of the Northwest likely will remain rural for many years to come. Three years ago when the Laboratory was established, the report of the Five-State Task Force identified the plight of rural schools in the Northwest region as one of four educational priorities. NWREL's Small Schools Program addresses itself to one of these.

Four crucial needs were cited by the report as having special relevancy to rural education in the Northwest. They were:

1. The lack of adequately trained teaching personnel, usually characterized by a high turnover rate, and the lack of inservice opportunities
2. A narrow and frequently out-of-date curriculum offering
3. The low aspiration level of rural students

4. The economic and cultural deprivation existent in these geographically isolated environments ("rural" and "poverty" are nearly synonymous)

NWREL has developed several activities focused on these problems of isolated schools. For example the Intercultural program has been working with the creation, development and testing of readers for Alaskan natives and Indians. Built upon the graphoneme concept, they are being field tested in seventeen villages in Alaska. This testing involves the Bureau of Indian Affairs, the Indian Tribal Councils, the State Department of Education, the University of Alaska anthropologists, linguists and others. Time, however, does not allow me to speak in detail on the readers.

For the purpose of this presentation, one activity has been chosen as an example from among the six that comprise the Small Schools Program. It should illustrate how NWREL has interpreted its task of forming a bridge between research knowledge and classroom practice. There are many roles that must be played by Laboratory personnel in dealing with each agency or group in an attempt to bring about quality changes in the educational establishment.

The use of this example, however, is not intended to suggest that rural educational problems have been solved in the Northwest. They have not! *Patterns In Arithmetic* (PIA), however, does hold promise for making some improvements in the rural environment. The Laboratory has been using PIA to improve education in rural isolated schools. The sequence was developed at the Research and Development Center, University of Wisconsin.

PIA is a modern elementary mathematics program for grades 1–6 consisting of videotape TV lessons, teachers' manuals and pupil exercise booklets (based on the work of Van Engen).* The 15-minute lessons introduce new concepts, review previously covered concepts and skills, and provide motivation toward the study and understanding of arithmetic. Grade 1 consists of 32 lessons, approximately 1 per week; grade 2 has 48 lessons with 3 lessons every 2 weeks; grades 3–6 consists of 2 lessons per week. The series attempts to update teachers in modern math principles and teaching methods while students are being taught basic arithmetic concepts. Original field testing in 1966–67 included 9,000 students in Wisconsin and Alabama. This school year over 138,000 students throughout the nation are using PIA.

* Van Engen, Henry, Assoc. Director, Development

(Educational Testing Service participated in this program by developing the midterm and final tests.)

A careful review of PIA by rural educators confirmed the ideas of NWREL staff that it offered promise in helping rural elementary teachers. In such cases, the Laboratory plays several roles in introducing a new approach to regional educators.

NEGOTIATOR Appropriate arrangements were made with the Wisconsin R & D Center and with the National Instructional Television Center to convert the two-inch videotape recorder tapes to one-inch reels, enabling them to be used on less expensive, portable video equipment. Now, PIA can be used in rural areas where educational television stations are nonexistent.

SALESMAN Consistent with the NWREL philosophy of working with and through existing organizations, approaches were made to three State Departments of Education in the Northwest region. The potential of PIA was explained and assurances were given that its use would not infringe upon the State Department's service and/or its supervising function.

CATALYST In addition, the Laboratory felt that obtaining a commitment from these State Departments of Education to engage in testing, monitoring and evaluating new products like PIA would facilitate a new role for each of these agencies.

EDUCATOR Subsequent conferences were successful with key State Department personnel in the three states. Agreements were reached whereby NWREL would supply the PIA materials and videotape equipment. The Laboratory also would provide the evaluation procedures to be employed. The State Departments agreed to select the sites, monitor the action, collect the data and appoint a person to coordinate these efforts.

TEACHER Conferences in each state were actually teaching-learning sessions devoted to understanding change processes and evaluation strategies. For example, some of the factors considered were:

1. The need to overcome the suspicion of rural teachers, administrators and community leaders of "outsiders who dress, act and talk differently"
2. The wariness which rural female teachers have of modern technological equipment, e.g., a survey of 16 schools found 1 movie projector

3. The need to engage in collecting data and information to evaluate the use of the materials and determine the problems that developed
4. The roles to be played by State Department personnel, the county superintendent, the local administrator and teachers, e.g., over-supervision was a worry
5. The determination of necessary data and gathering techniques.

DEMONSTRATOR At this point, what might loosely be called a remote control operation began. NWREL demonstrated the use of the materials and the equipment to State Department personnel. They, in turn, demonstrated them to local school boards, administrators, teachers and students. Laboratory personnel worked through the problems of collecting data and information regarding the use of PIA in the local district and suggested alternative methods of evaluation. Forms were constructed as guidelines for site visitations. Interview forms also were devised to obtain the reactions of pupils, teachers and administrators. In short, a formative evaluation design was developed that met the requirements of both the State Departments of Education and the Laboratory.

FOLLOWUP NWREL field staff kept in close contact with State Department personnel during the school year. They also visited each of the sites as part of the general plan. On May 15, 1969, State Department of Education personnel from the three states reviewed the preliminary report and suggested modifications. Later that month, a feedback session was scheduled with the authors. Finally, NWREL personnel have been preparing the report for the University of Wisconsin Center regarding the experiences of using PIA in rural, isolated elementary schools in the Northwest region. Center personnel then must decide the extent to which they can modify the PIA materials to more nearly meet the needs of these teachers and pupils. Together with the State Departments, the Laboratory is planning the work for the 1969–70 school year. Naturally, these plans depend upon the modification decisions by the staff of the Wisconsin Research and Development Center staff.

PRELIMINARY RESULTS FROM FIELD TRIALS

Geographically, this is a far-flung operation which covers one remote school in Alaska, two rural schools in Montana and three in Idaho. Year-end test results are still being prepared. Preliminary data have revealed:

1. State Department personnel initially committed to the project from all three states were all positive concerning their
   a. New role in innovation and change
   b. Attitude toward *Patterns in Arithmetic* as a useful tool in upgrading the skills of rural teachers
   c. Willingness to continue and if possible, expand the program next year
2. From the results tabulated thus far, pupil achievement compared favorably with pilot tests conducted by University of Wisconsin evaluators. There, 70 per cent of the 9,000 students performed above the 50th percentile on standardized achievement tests.
3. Rural isolated teachers with a minimal orientation to the materials were able to use them successfully. In addition, all of them have overcome their early fears of using the videotape recorder.
4. All teachers using *Patterns in Arithmetic* materials indicated their own understanding of mathematics had grown measurably during the year. Appreciation was expressed for the "whys" of arithmetic being so carefully explained by the TV teacher.
5. In a selected sample, interviews were conducted with first and second grade students from classes where PIA was used. They revealed eight out of ten chose arithmetic as their favorite subject in school as compared to a similar group of nonusers, the majority of whom indicated reading was their favorite area of study.
6. Improvements suggested by rural teachers who used *Patterns in Arithmetic* included:
   a. Keying the Teacher's Manual to the video presentations and the student worksheets
   b. Identifying the mathematical concepts to be attained in each video presentation
   c. Preparing diagnostic tests to enable easier entry into the materials by the students
   d. Indicating in the student workbook the skills needed to perform specific lessons satisfactorily
   e. Providing more feedback information to the teachers on student performance following the television presentations and the use of the worksheets
   f. Including discussion ideas and helps in the Teacher's Manual to expand the specific concepts presented

g. Improving the quality of specific film clips for grades 1 and 2

Some interesting questions have arisen from this activity. For example:

1. What new role expectancies are raised for the teacher? *Patterns In Arithmetic* presents the basic content; the classroom teacher becomes supplementary. Where do teachers and paraprofessionals fit into materials such as PIA?
2. At what point in the development process is it necessary to introduce the potential user? Should he enter the process early, and how involved in its development should he become?
3. Should new student materials being developed consciously build in the inservice factor for teachers, thus updating their skills?
4. Will this involvement process result in the establishment of a network that may be used successfully in the future for adaption/adoption of other innovations?

Laboratory staff already have engaged in discussions with researchers at R & D Centers and Universities to seek help in gaining new insights into these and other concerns growing out of this experience.

Before concluding, a few general remarks seem appropriate regarding the Laboratory setting and the Northwest Regional Educational Laboratory specifically.

As you are aware, laboratories were created to serve as a bridge between what is *known* and what is *practiced* in education. Their aim is to speed the movement of quality improvements within the educational establishment. This unique idea demonstrates the principle of creative federalism by returning federal tax funds to the local level to be used on local needs as determined by local policy makers. Each laboratory has the freedom to create its own programs and to determine its own strategies for attacking these educational problems.

At the Northwest Regional Educational Laboratory there is a fundamental belief that education can be better—in content, procedures, technology and organization. NWREL is clearly an advocate for improving education, for innovation and for change. Its primary mission is developing new and tested alternatives for educators. Fundamental to its strategy for change, NWREL relies heavily upon involvement of the institutions, organizations, associations and individuals with whom it works. *Products* could be developed more quickly if concentrated upon in an environment isolated from the setting where they are to be used. But the Laboratory believes

change will come about more quickly and will be more permanent if those involved are active participants in the development process.

Involvement of potential users/adopters begins with the selection of an activity from among available alternatives: it may continue through prototype development, field testing, evaluation, demonstration and final adoption. A selected activity can enter this process at any point, cycling back and forth as needed. In reality this becomes more of a circle of development than a straight line progression.

The NWREL commitment to a developmental strategy demands a wide variety of roles be played by its personnel. It is a constant concern of Laboratory staff members that no role be assumed which properly belongs to others. The Laboratory has been designed to serve as an extension or to complement the roles played by other agencies or individuals.

Returning to the five-step planning model initially introduced:

Step 1. Rural isolated elementary teachers often lack the formal training and/or inservice opportunities needed to update skills in mathematics. *They hurt!*

The students taught by teachers lacking formal training are geographically isolated and often economically deprived. Many are forced to accept a narrowly oriented curriculum which is often woefully out of date. The children don't compete. *They hurt!*

Step 2. The low economic capability of many rural schools prevents them from paying adequate salaries to attract well qualified teachers. Geographic isolation often prevents community leaders, school boards and administrators from realizing the educational offerings of their school are not adequate to compete with their urban and suburban counterparts. The meager fare offered students in a stilted environment affects the aspiration level of students. (In one community, no child in three generations of one family had completed high school.) That is *why they hurt!*

Step 3. The PIA program offered rural isolated elementary teachers an inservice opportunity to update their mathematic skills. Students, likewise, were afforded the opportunity to participate in a high quality teaching experience and to learn new, modern mathematics, thus enhancing their chances of competing. They both received *the aspirin!*

Step 4. The field trials of PIA in five isolated areas of the Northwest showed deficiencies in the materials and the equipment used. Mathematics is only one of the many subject areas requiring attention. *No, the aspirin is not as big as the headache!*

Step 5. The feedback sessions with Center personnel will give Laboratory staff an opportunity to describe and enumerate the different headaches caused by the adoption of PIA in rural isolated schools. Hopefully, a better aspirin will be developed. At least we hope for one that may reduce the mathematical headaches for rural elementary teachers and pupils!

Before concluding, a few general remarks seem appropriate regarding the laboratory setting and the Northwest Laboratory specifically.

As indicated earlier today, regional laboratories are very, very, young organizations and will be celebrating their third anniversary in June. As struggling new agencies, they have been required to mature rapidly. A 25 per cent mortality rate during infancy has been high. Perhaps it may increase.

As you are aware, laboratories were created to serve as a bridge between what is known and what is practiced and to speed quality improvements within the educational establishment. Each laboratory has had great freedom in creating its own programs and in determining its own strategies for attacking these educational problems.

At the Northwest Laboratory there is a fundamental belief that education can be better—in content, procedures, technology and organization. The Northwest Laboratory is clearly an advocate for improving education, for innovation and for change. Its primary mission is developing new and tested alternatives for educators. Fundamental to its strategy for change, Northwest Laboratory relies heavily upon the involvement of institutions, organizations, associations and individuals with whom it works. Products could be developed more quickly if concentrated upon in an environment isolated from the setting where they are to be used. The Laboratory believes that change will come about more quickly and will be more permanent if those who are the focus of change are active participants in the development process.

Involvement of potential users and adopters begins with the selection of an activity from among available alternatives. It may continue through prototype development, field testing, evaluation, demonstration and final adoption. A selected activity can enter this process at any point, cycling back and forth as needed. In reality this becomes more of a circle of development than a straight line progression.

The Laboratory's commitment to a developmental strategy demands that a wide variety of roles be played by its personnel. The example used earlier illustrated some of these roles well; for example, advocate, catalyst, risk taker and sharer, teacher, learner, and often

patient listener. It is a constant concern of Laboratory staff members that no role be assumed that properly belongs to others. The Laboratory should serve as an extension or complement to the roles played by other agencies or individuals.

As a member of such an organization for the past three years, interesting challenges have been presented. To mention a few, I have enjoyed being actively and intimately involved in here-and-now educational problems; constantly searching for new and better alternatives; and trying them out with built-in feedback mechanisms that report the selected materials and strategies that result in modifying future actions. In short, it is a data-based organization.

I also have enjoyed the possibility of being able to plan sustained long-term efforts. Contrary to some people's belief that we are in business for only one year at a time, we believe we are going to be around for a long time. We look forward to being able to sustain long-term efforts which have the promise of accumulating a critical mass of experience, information and data upon specific problems or situations.

The third thing, to be able to marshall knowledgeable experts to aid in determining priorities and strategies of attack on chosen problems, evaluate these efforts, to redefine or refocus as needed, has also been an excitement to me.

CONCLUSION

My purpose in this presentation of one specific NWREL activity was to:

1. Demonstrate, by example, how a Laboratory is attempting to bridge the gap between research knowledge and educational practice.
2. Illustrate the intricacies involved in working with many agencies to improve rural education.
3. Show the many roles that must be played by Laboratory personnel.

I hope I have also been able to convey my enthusiasm and belief in the laboratory concept as a new means of making educational improvements.

Conference List

Fred C. Adams, *The Know and Care Center, San Mateo Union High School District*, San Mateo, California
John D. Adams, Jr., *Burlingame High School*, Burlingame, California
A. M. Akers, *Sierra College*, Rocklin, California
Helene S. Aldez, *Reno High School*, Reno, Nevada
Marvin Alkin, *Center for the Study of Evaluation of Instructional Programs, University of California*, Los Angeles, California
Gerald J. Alves, *Sonoma State College*, Rohnert Park, California
James R. Anderson, *San Jose Unified School District*, San Jose, California
Lorine A. Aughinbaugh, *American River College*, Sacramento, California
Garnet L. Austin, *Livermore Valley Joint Unified School District*, Livermore, California
Dayton Axtell, *Merritt College*, Oakland, California
Orville R. Bailey, *Scappoose School District*, Scappoose, Oregon
J. T. Ball, *Chabot College*, Hayward, California
Valerie Banks, *Mills College*, Oakland, California
J. M. Bardellini, *Mount Diablo Unified School District*, Concord, California
Jerome Beamish, *Stanislaus State College*, Turlock, California
Joseph A. Benedict, *San Lorenzo Valley Unified School District*, Ben Lomond, California
Orrin Berg, *Pacific College*, Fresno, California
Mary L. Betts, *Palo Alto Unified School District*, Palo Alto, California
John Bianchini, *Educational Testing Service*, Berkeley, California
Col. Lee C. Black, *Preparatory School Academy, U.S.A.F. Academy*, Colorado
Ken Blaker, *University of Santa Clara*, Santa Clara, California
Don F. Blood, *Western Washington State College*, Bellingham, Washington
Walter R. Borg, *Far West Regional Laboratory*, Berkeley, California
Robert L. Bovinette, *Occidental College*, Los Angeles, California
Guy H. Browning, *Stanford University*, Stanford, California
Katherine I. Brownlie, *Educational Testing Service*, Berkeley, California
Donald Boysen, *Saratoga High School*, Saratoga, California
Robert Bush, *Castro Valley Unified School District*, Castro Valley, California

Gerald G. Cain, *Southern Methodist University*, Dallas, Texas
Robert G. Cameron, *CEEB Western Regional Office*, Palo Alto, California
Patrick C. Cannon, *South San Francisco Unified School District*, South San Francisco, California
Richard Carey, *Palo Alto Unified School District*, Palo Alto, California
William C. Carey, *Pacific Grove Unified School District*, Pacific Grove, California
Dale Carlson, *State Department of Education*, Sacramento, California
Abraham Carp, *Educational Testing Service*, Berkeley, California
A. Edward Chally, *Davis Joint Unified School District*, Davis, California
C. Douglas Chretien, *University of California*, Berkeley, California
Paul Christensen, *University of California*, Berkeley, California
Leonard J. Clark, *Modesto City Schools*, Modesto, California
Margaret W. Clark, *San Jose State College*, San Jose, California
W. H. Clinkenbeard, *Los Angeles County Schools*, Los Angeles, California
Randall A. Cognetta, *San Mateo City School District*, San Mateo, California
Reginald Corder, *Educational Testing Service*, Berkeley, California
Myles M. Cosgrove, *Abraham Lincoln High School*, San Francisco, California
Charles D. Cox, *Simpson Bible College*, San Francisco, California
Robert Crawford, *Pacific Grove Unified School District*, Pacific Grove, California
Patricia Cross, *Educational Testing Service*, Berkeley, California
Roger W. Cummings, *San Francisco State College*, San Francisco, California
James R. Cunningham, *Humboldt State College*, Arcata, California
Robert Joseph Darling, *State Department of Education*, Sacramento, California
Art Dambacher, *Berkeley Unified School District*, Berkeley, California
Anne J. Davis, *U.C. Medical Center, School of Nursing*, San Francisco, California
Junius A. Davis, *Educational Testing Service*, Durham, North Carolina
Judy Dawson, *Riverside Unified School District*, Riverside, California
Joe De Wees, *Modesto City Schools*, Modesto, California
Jackson C. Dickson, *San Lorenzo Unified School District*, San Lorenzo, California
Adelle DiGiorgio, *CEEB Western Regional Office*, Palo Alto, California
Valeria Dotterrer, *East Los Angeles College*, Los Angeles, California
Carolyn Doty, *School Testing Service*, Berkeley, California
Sidney Duddy, *Tamalpais High School*, Mill Valley, California
Glenn W. Durflinger, *University of California*, Santa Barbara, California
Arnold J. Edman, *West Valley College*, Campbell, California
Kenneth C. Edson, *Chico State College*, Chico, California
Catherine Farley, *Merritt College*, Oakland, California
Theodore T. Fong, *Alum Rock Elementary School District*, San Jose, California
Richard Fortna, *Educational Testing Service*, Berkeley, California

Theodore E. Foster (Mrs.), San Francisco, California
Carolyn M. Fowle, *Lodi Unified School District*, Lodi, California
John L. Fritz, *Watsonville High School*, Watsonville, California
George Fujinaga, *Educational Testing Service*, Berkeley, California
Robert J. Gilbride, *Terra Linda High School*, San Rafael, California
Fred Gillette, *San Francisco State College*, San Francisco, California
Donald P. Glaser, *California Teachers Association*, Burlingame, California
Gordon R. Graves, *Fresno City Unified School District*, Fresno, California
Frank H. Gurley, *Stanislaus County Schools*, Modesto, California
Howard R. Haering, *San Bernardino County Schools*, San Bernardino, California
Jack Hamilton, *American Institute for Research*, Palo Alto, California
Gerald Hanson, *Sunnyvale High School*, Sunnyvale, California
Lucille Harris, *Orange County Schools Office*, Santa Ana, California
Diane Harsh, *San Francisco Art Institute*, San Francisco, California
J. Richard Harsh, *Educational Testing Service*, Los Angeles, California
Carolyn S. Hartsough, *University of California*, Berkeley, California
Einar Haugen, *Golden Gate Academy*, Oakland, California
Harrison F. Heath, *San Jose State College*, San Jose, California
Nancy Heinz, *California State College*, Hayward, California
John Helmick, *Educational Testing Service*, Berkeley, California
John K. Hemphill, *Far West Regional Laboratory*, Berkeley, California
Jean Henderson, *California College of Arts and Crafts*, Oakland, California
Walter S. Hertzog, *East Los Angeles College*, Los Angeles, California
Herbert Hogan, *La Verne College*, La Verne, California
Donald E. Hood, *Educational Testing Service*, Austin, Texas
May I. Howe, *Reno High School*, Reno, Nevada
Thomas F. Humiston, *City College of San Francisco*, San Francisco, California
Alice J. Irby, *Educational Testing Service*, Princeton, New Jersey
William Jackman, *Holmes Junior High School*, Davis, California
Walter Jackson, *Grant Joint Union High School*, Sacramento, California
Frances Jamieson, *Diagnostic School for Neurologically Handicapped Children*, San Francisco, California
Gurbachan Jauhal, *Morgan Hill Unified School District*, Morgan Hill, California
Iva Jex, *School Testing Service*, Berkeley, California
Reynold B. Johnson, *IBM Corporation*, Los Gatos, California
Carol Y. Jones, *Educational Testing Service*, Berkeley, California
Nell Jones, *Educational Testing Service*, Berkeley, California
Ray Jongeward, *Northwest Regional Education Laboratory*, Portland, Oregon
Keith Kading, *Folsom-Cordova Unified School District*, Folsom, California
Warren Kallenbach, *San Jose State College*, San Jose, California
Louis Kaplan, *Torrance School District*, Del Amo, California
Mrs. Louis Kaplan, *Torrance School District*, Del Amo, California
Betsy Kellman, *Katherine Burke School*, San Francisco, California
Donald L. Kester, *Chabot College*, Hayward, California

Jack W. Kirby, *San Rafael City Schools*, San Rafael, California
Stephen Klein, *University of California at Los Angeles*, Los Angeles, California
Santelia Knight, *Educational Testing Service*, Berkeley, California
Richard L. Knowles, *Santa Clara Unified School District*, Santa Clara, California
John W. Lackey, *Southern California College*, Costa Mesa, California
Nadine M. Lambert, *University of California*, Berkeley, California
Robert E. Lambert, *Educational Testing Service*, Berkeley, California
William H. Landis, *Contra Costa County Department of Education*, Pleasant Hill, California
W. Timothy Lange, *Fullerton Elementary School District*, Fullerton, California
Alexander I. Law, *State Department of Education*, Sacramento, California
Lawrence Lew, *San Francisco Unified School District*, San Francisco, California
Charles R. Lewin, *San Luis Coastal Unified School District*, San Luis Obispo, California
John H. Lewis, *Cubberley Senior High School*, Palo Alto, California
David Logothetti, *University of Santa Clara*, Santa Clara, California
Bruce Lowrey, *State Department of Education*, Sacramento, California
Sister Juliana Lucey, *College of the Holy Names*, Oakland, California
Mary Jane McMahon, *Tuolumne County Schools*, Sonora, California
Thomas F. MacMillan, *Napa Community College*, Napa, California
Patricia Mannion, *Educational Testing Service*, Berkeley, California
Herschel T. Manuel, *Guidance Testing Association*, Austin, Texas
Kenneth K. Martin, *University of Alaska*, College, Alaska
Jane E. Matson, *California State College*, Los Angeles, California
A. John Mattila, *Sunnyvale School District*, Sunnyvale, California
Chester Maul, *San Mateo County Schools*, Redwood City, California
Frances Mettler, *Palo Alto Unified School District*, Palo Alto, California
Edie A. Moon, *Sonoma State College*, Rohnert Park, California
Gary Moore, *Santa Clara Unified School District*, Santa Clara, California
Donald G. Mortensen, *California State College*, Los Angeles, California
Howard K. Morton, *California State College*, Fullerton, California
Virginia Murdoff, *Napa Community College*, Napa, California
Ruth L. Murray, *San Francisco State College*, San Francisco, California
Oberlin B. Nereson, *Foothill College*, Los Altos, California
Arthur Neumann, *Educational Testing Service*, Berkeley, California
Richard Newman, *Tamalpais High School*, Mill Valley, California
Sheila G. Newman, *Los Angeles Unified School District*, Los Angeles, California
Warren Newman, *Burbank Unified School District*, Burbank, California
Gloria Nicoll, *Redwood City School District*, Redwood City, California
Richard E. Nunes, *Contra Costa County Schools*, Pleasant Hill, California
Phillip W. Oakes, *San Juan Unified School District*, Carmichael, California
Laurell E. Olson, *Williams High School*, Williams, California
Sonja Orescan, *Educational Testing Service*, Berkeley, California

Mildred Ostich, *Seventh Day Adventist Schools*, Oakland, California
Mary H. Owen, *Educational Testing Service*, Berkeley, California
Bruce L. Paulson, *Mt. San Antonio College*, Walnut, California
Casper F. Paulson, Jr., *Oregon College of Education*, Monmouth, Oregon
Frank Peluso, *San Jose State College*, San Jose, California
Lyman Peterson, *River Delta Unified School District*, Rio Vista, California
Richard Peterson, *Educational Testing Service*, Berkeley, California
Bernadine Plamondon, *Army Education Center*, San Francisco, California
George V. Porter, Jr., *United States Air Force Academy*, Colorado
Jack Price, *Tuolumne County Schools*, Sonora, California
Mabel C. Purl, *Riverside Unified Schools*, Riverside, California
Willamay Pym, *Shoreline Community College*, Seattle, Washington
Alvin B. Quall, *Whitworth College*, Spokane, Washington
Richard M. Regedal, *Kern High School District*, Bakersfield, California
Allen Reitzenstein, *Jackson County Intermediate Education District*, Medford, Oregon
Richard H. Reyes, *San Francisco State College*, San Francisco, California
Joseph P. Rice, *California State Department of Education*, Sacramento, California
Malby Roberts, *Cabrillo College*, Aptos, California
David J. Robertson, *Yuba County Schools Office*, Marysville, California
Wilson S. Robinson, *Golden Gate College*, San Francisco, California
Thomas B. Rodgers, *Educational Testing Service*, Berkeley, California
Ben Romine, *Regional Educational Laboratory*, Durham, North Carolina
Charles Rothaermel, *California High School*, Whittier, California
Dorothy Roy, *Educational Testing Service*, Berkeley, California
Glenn Royer, *Napa Valley Unified School District*, Yountville, California
Robert Sayette, *East Los Angeles College*, Los Angeles, California
Wade Schroeder, *Campbell Union High School District*, Campbell, California
Jack B. Selbig, *University of Nevada*, Reno, Nevada
Dwight T. Shafer, *San Jose State College*, San Jose, California
Marion F. Shaycoft, *American Institutes for Research*, Palo Alto, California
Thomas Shellhammer, *California State Department of Education*, Sacramento, California
Ralph P. Sherlock, *Merced Union High School District*, Merced, California
Arthur W. Sherman, *Sacramento State College*, Sacramento, California
William H. Sherrill, *University of California*, Berkeley, California
Edmond O. Shinn, *College of San Mateo*, San Mateo, California
William M. Slaichert, *University of Denver*, Denver, Colorado
Dorothy Soeberg, *Riverside County Schools*, San Clemente, California
Charles L. Solbach, *Watsonville High School*, Watsonville, California
Bette Spagel, *Mills College*, Oakland, California
Willem Spits, *Educational Testing Service*, Berkeley, California
Curtis R. Stafford, *San Jose State College*, San Jose, California
Vivian Stallkamp, *Campbell Union High School District*, Campbell, California
James E. Sudyk, *Gunn Senior High School*, Palo Alto, California

Lee Ann Taylor, *San Francisco State College*, San Francisco, California
George Temp, *Educational Testing Service*, Berkeley, California
Alice C. Thompson, *California State College*, Los Angeles, California
Ivor J. Thomas, *Educational Testing Service*, Los Angeles, California
Edward R. Tibby, *Educational Testing Service*, Berkeley, California
Pauline Toschi, *College of Marin*, Kentfield, California
Vernon E. Trimble, *Saratoga High School*, Saratoga, California
LeRoy J. Tuscher, *Campbell Union High School District*, Campbell, California
Barbara Varenhorst, *Palo Alto Unified School District*, Palo Alto, California
Jack Vassar, *San Leandro High School*, San Leandro, California
Paul Verdone, *Sacramento State College*, Sacramento, California
Douglas R. Walker, *Stanford University*, Stanford, California
Harvey R. Wall, *Mt. Diablo Unified School District*, Concord, California
Jim Walter, *Oregon College of Education*, Monmouth, Oregon
Margaret C. Wang, *University of Pittsburgh*, Pittsburgh, Pennsylvania
Jonathan Warren, *Educational Testing Service*, Berkeley, California
Richard W. Watkins, *Far West Laboratory for Educational Research and Development*, Berkeley, California
Merrill Joseph Weijola, *California State College*, Long Beach, California
John A. Wennergren, *Gonzales Union High School*, Gonzales, California
Charles Wheeler, *Fresno State College*, Fresno, California
Robert G. Wolstoncroft, *Loma Linda University*, Riverside, California
Floyd G. Wood, *Loma Linda University*, Riverside, California
Kenneth A. Wood, *California Community Colleges*, Sacramento, California
James C. Young, *Kern High School District*, Bakersfield, California
Leil L. Young, *College of Notre Dame*, Belmont, California